

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 12:26:02 2000; MasPar time 2.83 Seconds
Tabular output not generated. 75.274 Million cell updates/sec

Title: >US-09-376-430-2
Description: (48-56) from US09376430A.pep (10 of 25)
Perfect Score: 58
Sequence: 1 ASKYSRTNL 9

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:geneseq

Statistics: Mean 16.090; Variance 38.671; scale 0.416

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	46	79.3	802	1 W60570	Hydronephrosis gene (H	6.22e+01
2	44	75.9	789	1 R47519	EHV-4 ribonucleotide r	1.13e+02
3	43	74.1	255	1 W01204	Serine protease PfSP7-	1.51e+02
4	43	74.1	342	1 W55938	Human par-4 protein.	1.51e+02
5	43	74.1	400	1 P70654	Sequence encoded by g	1.51e+02
6	43	74.1	694	1 R14666	Truncated HSVB polyep	1.51e+02
7	43	74.1	903	1 P70426	Recombinant herpes sim	1.51e+02
8	43	74.1	903	1 P60244	Herpes simplex virus g	1.51e+02
9	43	74.1	903	1 P50312	Herpes simplex virus 1	1.51e+02
10	43	74.1	903	1 P71135	Herpes simplex virus 1	1.51e+02
11	43	74.1	904	1 R41779	HSVGB polyepptide.	1.51e+02
12	43	74.1	904	1 R41779	Glycoprotein B (gB2).	1.51e+02
13	43	74.1	904	1 W34553	Herpes simplex virus t	1.51e+02
14	43	74.1	904	1 W00375	HSV-1 glycoprotein B.	1.51e+02
15	43	74.1	905	1 P80915	Sequence of Herpes sim	1.51e+02
16	43	74.1	973	1 P70769	Glycoprotein B of herp	1.51e+02
17	43	74.1	973	1 R14680	HSV surface antigens g	1.51e+02
18	43	74.1	1423	1 W71367	Death associated prote	1.51e+02
19	43	74.1	1423	1 R74205	Human death associated	1.51e+02
20	42	72.4	248	1 W20921	H. pylori derived prot	2.03e+02
21	42	72.4	326	1 W36985	Haemophilus influenzae	2.03e+02
22	42	72.4	477	1 W35890	E. coli glycogen synth	2.03e+02
23	42	72.4	477	1 R25462	glgA.	2.03e+02

24	42	72.4	798	1 R85879	WD-40 domain-contg. TU	2.03e+02
25	42	72.4	798	1 W33634	Yeast transcriptional	2.03e+02
26	42	72.4	798	1 R91304	TATA box binding prote	2.03e+02
27	42	72.4	798	1 R85883	WD-40 domain-contg. ye	2.03e+02
28	41	70.7	330	1 W86252	Maize PTE mature prote	2.70e+02
29	41	70.7	332	1 W86253	Amino acid sequence of	2.70e+02
30	41	70.7	419	1 W86251	Maize PTE enzyme.	2.70e+02
31	41	70.7	422	1 W71639	Omega-cyclohexane fatt	2.70e+02
32	41	70.7	585	1 R96246	Malic enzyme #1.	2.70e+02
33	41	70.7	638	1 R70096	Malic acid enzyme isol	2.70e+02
34	41	70.7	2089	1 W08333	Cyclocella cryptica ac	2.70e+02
35	40	69.0	32	1 R15809	Calcitonin deriv. (10a	3.60e+02
36	40	69.0	375	1 R74142	Mango class I thioeste	3.60e+02
37	40	69.0	375	1 W44332	Class I acyl-ACP thioe	3.60e+02
38	40	69.0	413	1 Y07057	Renal cancer associate	3.60e+02
39	40	69.0	447	1 W13381	Sesame omega-3 aliphat	3.60e+02
40	40	69.0	559	1 R05427	Circumsporozoite (CS)-	3.60e+02
41	40	69.0	607	1 W95601	Human secretory immuno	3.60e+02
42	40	69.0	657	1 R28964	Notch hN3k full length	3.60e+02
43	40	69.0	746	1 W03178	Human poly-immunoglobu	3.60e+02
44	40	69.0	1078	1 R28963	Notch hN3k full length	3.60e+02
45	40	69.0	2471	1 Y06816	Human Notch2 (humN2) p	3.60e+02

ALIGNMENTS

RESULT 1						
ID	W60570	standard; Protein; 802 AA.				
AC	W60570;					
DE	04-SEP-1998 (first entry)					
DR	Hydronephrosis gene (HNG) product sequence.					
KW	Hydronephrosis gene; HNG gene; USF2 gene; renal disease; renal aplasia;					
KW	vesical-ureteral reflux; pelvi-ureteral junction obstruction;					
KW	multicystic renal dysplasia; renal agenesis; hydronephrosis;					
OS	Von Mayer-Rokitansky-Kuester disorder; bilid ureter.					
FS	Homo sapiens.					
FT	Key	Location/Qualifiers				
FT	Peptide	200..204				
FT	Peptide	/note= "putative nuclear localisation signal"				
FT	Peptide	201..205				
FT	Peptide	/note= "putative nuclear localisation signal"				
FT	Peptide	268..271				
FT	Peptide	/note= "putative nuclear localisation signal"				
FT	Domain	378..500				
FT	Domain	/note= "putative transcriptional activating domain"				
PN	W09815650-A2.					
PD	16-APR-1998.					
PF	09-OCT-1997; E05583.					
PR	09-OCT-1996; EP-202820.					
PA	(VLAA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.					
PI	Fryns JPGJ, Groenen PMA, Van De Ven WJM;					
DR	WPI: 98-240833/21.					
DR	N-PDB; V18098.					
PT	Hydronephrosis gene - useful to treat or diagnose renal diseases and					
PT	disorders, e.g. vesical-ureteral reflux, pelvi-ureteral junction					
PT	obstruction, multicystic renal dysplasia or renal agenesis					
PS	Disclosure: Fig 9: 73pp; English.					
CC	This represents a hydronephrosis gene (HNG) product. The HNG gene can be					
CC	used as a starting point to design suitable compounds or techniques for					
CC	the treatment of renal diseases or disorders, or nucleotide probes for					
CC	diagnosing cells involved in renal diseases or disorders. A protein or a					
CC	fragment encoded by HNG gene can be used as a starting point for					
CC	preparing suitable antibodies for diagnosing cells involved in renal					
CC	diseases and disorders. The products and method can be used to treat or					
CC	diagnose renal diseases and disorders selected from vesical-ureteral					
CC	reflux, uni- or bilateral pelvi-ureteral junction obstruction, multicystic					
CC	renal dysplasia, renal agenesis, renal aplasia, hydronephrosis,					
CC	Von Mayer-Rokitansky-Kuester disorder and bilid ureter.					
SQ	Sequence 802 AA;					

Query Match 79.3%; Score 46; DB 1; Length 802;
Best Local Similarity 62.5%; Pred. No. 6.22e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 675 SRYTRANL 682
I:|:|:|
QY 49 SKYSRTNL 56

RESULT 2
ID R47519 standard; Protein; 789 AA.

AC R47519;
DT 08-JUL-1994 (first entry)
DE EHV-4 ribonucleotide reductase subunit RR1 (wild-type).
KW EHV-4; ribonucleotide reductase subunit; RR1; insertion region;
KW mutein; mutant; variant; vaccine; attenuated.
OS Equine herpesvirus 4.
PN W09400587-A.
PD 06-JAN-1994.
PF 29-JUN-1993; G01355.
PR 30-JUN-1992; GB-013882.
PA (EQU1-) EQUINE VIROLOGY RES FOUND.
PA (UNIU) UNIV GLASGOW.
PI Onions DE, Riggiolo M;
DR WPI; 94-026223/03.
DR N-PSDB: Q53996.
PT Equine herpes virus-4 mutated in ribonucleotide reductase gene -
PT used in vaccine opt. contg. DNA encoding second antigen
PS Claim 1; Fig 6; 37pp; English.
CC The wild-type gene coding for native EHV-4 ribonucleotide reductase
CC subunits 1 and 2 (Q53996) was isolated from cultures of equine
CC dermal cells (NBL-6) which had been experimentally infected with
CC EHV-4 strain 1942. Mutants of EHV-4 are claimed which comprise a
CC mutation in the RR1 or RR2 coding regions. The mutation is pref.
CC caused by insertion of a heterologous coding sequence, esp. one that
CC codes for a different equine pathogen such as EHV-1, Streptococcus
CC equi, equine rotavirus or encephalitis virus.
SQ Sequence 789 AA;

Query Match 75.9%; Score 44; DB 1; Length 789;
Best Local Similarity 75.0%; Pred. No. 1.13e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 575 SKYSRGEL 582
|:|:|:|:|
QY 49 SKYSRTNL 56

RESULT 3

ID W01204 standard; Protein; 255 AA.
AC W01204;
DT 16-MAR-1997 (first entry)
DE Serine protease PfSP7-255.
KW Flea; midgut; serine protease; PfSP7-255; vaccine;
KW domestic animal; infestation; insecticide; protease-inhibitor;
KW controlled release formulation; synergist.
OS Siphonaptera sp.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /note= "Signal peptide"
FT Peptide 2..98
FT /note= "Sequence W01190 (claim 71)"
FT Peptide 23..57
FT /note= "Sequence W01181 (claim 71)"
FT misc_difference 187
FT /note= "Amino acid not specified"
FT Peptide 152..154
FT /note= "Conserved GWA sequence"
FT Peptide 152..204
FT /note= "PfSP7-53 (W01172, claim 71)"
PN W09611706-A1.
PD 23-APR-1996.
PF 18-OCT-1995; U14442.
PR 18-OCT-1994; US-326773.
PR 07-JUN-1995; US-485455.
PR 07-JUN-1995; US-482130.

PR 07-JUN-1995; US-485443.
PR 07-JUN-1995; US-484211.
PA (PARA-) PARAVAX INC.
PA (HESK-) HESKA CORP.
PI Arfsten A, Dale B, Frank GR, Grieve RB, Heath A;
PI Hunter SW, Rushlow KE, Stiegler GL, Yamanaka M;
DR WPI; 96-221762/22.
DR N-PSDB; T40849.
PT DNA encoding Flea serine protease and aminopeptidase - useful in
PT vaccines to protect animals from flea infestation.
PS Claim 71; Page 181; 241pp; English.
CC This sequence (PfSP7-255) represents a flea midgut serine protease
CC encoded by the nfp7-894 gene, isolated from a flea cDNA library by
CC homology with conserved serine protease sequences (e.g. W01226).
CC The sequence contains a signal peptide and a conserved GWA
CC sequence. The sequence contains sequence W01172 (PfSP7-53), which
CC spans 2 conserved serine protease sequences, and N-terminal
CC sequences W01190 and W01181. The protein shows homology to a
CC hornet chymotrypsin-II and to collagenase. The protein may be used
CC in a vaccine for protection of domestic animals from flea
CC infestation, or in isolation of protease-inhibitors which may be
CC used in controlled release formulations to reduce the flea burden on
CC and around the animal. The inhibitors may be included in
CC insecticidal compositions to increase efficacy of other active
CC agents, by reducing proteolytic activity in the flea midgut.
SQ Sequence 255 AA;

Query Match 74.1%; Score 43; DB 1; Length 255;
Best Local Similarity 71.4%; Pred. No. 1.51e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 105 KYDRANL 111
|:|:|:|
QY 50 KYSRTNL 56

RESULT 4

ID W5958 standard; Protein; 342 AA.
AC W5958;
DT 12-AUG-1998 (first entry)
DE Human par-4 protein.
KW Human; ciao-1; par-4; WTL complex; Wilm's tumour; nephroblastoma;
KW binding protein; diagnosis; detection; tumour suppressor.
OS Homo sapiens.
PN W09813494-A2.
PD 02-APR-1998.
PF 26-SEP-1997; U17382.
PF 27-SEP-1996; US-028923.
PA (HARD) HARVARD COLLEGE.
PI Shi Y;
DR WPI; 98-230705/20.
DR N-PSDB; V26025.
PT Wilm's tumour WTL interacting proteins ciao-1 and par-4 - used to
PT develop products for modulating Wilm's tumour suppressor gene
PS Claim 9; Page 94-95; 122pp; English.
CC The present sequence represents par-4 which is a Wilm's tumour (WTL)
CC binding protein. Products of the invention or activity of the Wilm's tumour
CC suppressor gene, WTL. The products can be used for treating e.g. tumours,
CC endometriosis, ependymomas, for transplantation of organs, for treating
CC degenerative disorders of lung tissue caused by e.g. toxic injuries, as
CC well as inflammatory and degenerative processes induced by viral
CC infections. They can be used for treating neurological conditions
CC deriving from acute, subacute, or chronic injury to the nervous system,
CC including traumatic injury, chemical injury, and tumour-induced injury, chronic
CC together with infectious/inflammatory and tumour-induced injury, chronic
CC neurodegenerative diseases of the nervous system (e.g. Parkinson's
CC disease, Huntington's chorea, amyotrophic lateral sclerosis, or
CC spinocerebellar degenerations), and chronic immunological diseases of the
CC nervous system or affecting the nervous system (e.g. multiple sclerosis).
CC The products can also be used to treat neuromuscular disorders including
CC chronic atrophies, as well as other diseases which can manifest as
CC progressive bulbar palsies or spinal muscular atrophies. They can also

CC be used to rescue neurons in such conditions as CNS trauma, infarction,
 CC infection, metabolic disease, nutritional deficiency and after exposure
 CC to toxic agents.
 SQ Sequence 342 AA;

Query Match 74.1%; Score 43; DB 1; Length 342;
 Best Local Similarity 62.5%; Pred. No. 1.51e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 240 SRRYRTD 247
 :|||
 QY 48 ASKYRTN 55

RESULT 5
 ID P70654 standard; Protein; 400 AA.
 AC P70654;
 DT 16-APR-1991 (first entry)
 DE Sequence encoded by g B gene of Herpes simplex virus (HSVgB).
 KW Yeast expression vector.
 OS Herpes simplex virus.
 FH Key Location/Qualifiers
 FT peptide 1..30
 FT /label= signal peptide
 FT 31..400
 FT protein
 PN EP-216195-A.
 PD Q1-APR-1987.
 PF 29-AUG-1986; 112005.
 PR 30-AUG-1985; JP-192916.
 PA (KAGA) CEMO-SERO-THERAPEU.
 PI Nozaki C, Nakatake H, Kino Y, Eto T, Makizumi K, Ohtomo N;
 DR WPI; 87-087941/13.
 DR N-PSDB; N70101.
 PT Plasmid contg. herpes simplex virus gene - used for transforming
 PT yeast for producing HSV vaccines and diagnostic reagents
 PS Disclosure; Fig 1; 38pp; English.
 CC The HSVgB gene is used in the plasmid of the invention. The gB gene
 CC may be without the sequence encoding the N-terminal 30 or 83 AAs,
 CC including the signal peptide-encoding region.
 SQ Sequence 400 AA;

Query Match 74.1%; Score 43; DB 1; Length 400;
 Best Local Similarity 55.6%; Pred. No. 1.51e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 209 TAKYVRNRL 217
 :|||
 QY 48 ASKYRTN 56

RESULT 6
 ID R14666 standard; Protein; 694 AA.
 AC R14666;
 DT 23-JAN-1992 (first entry)
 DE Truncated HSVgB polypeptide.
 KW Vaccine; antigen.
 OS Herpes simplex virus.
 PN J03218397-A.
 PD 25-SEP-1991.
 PF 21-JUN-1990; 161448.
 PR 22-JUN-1989; JP-158238.
 PR 30-NOV-1989; JP-308941.
 PR 21-JUN-1990; JP-161448.
 PA (TAKE) TAKEDA CHEMICAL IND KK.
 DR WPI; 91-328397/45.
 DR N-PSDB; Q14479.
 PT HSVgB polypeptide(s) obtd. by recombinant DNA techniques -
 PT useful as vaccines against HSV and in diagnosis, can be produced
 PT cheaply and safely.
 PS Claim 1; Fig 7; 24pp; Japanese.
 CC The sequence was deduced from DNA and is that of a truncated HSVgB
 CC polypeptide. The recombinant protein can be used to prepare
 CC vaccines for prophylaxis of HSV infection and for use in diagnostic

CC kits.
 CC See also R14665.
 SQ Sequence 694 AA;

Query Match 74.1%; Score 43; DB 1; Length 694;
 Best Local Similarity 55.6%; Pred. No. 1.51e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 181 TAKYVRNRL 189
 :|||
 QY 48 ASKYRTN 56

RESULT 7
 ID P70426 standard; Protein; 903 AA.
 AC P70426;
 DT 30-APR-1991 (first entry)
 DE Recombinant herpes simplex virus.
 KW Herpes simplex virus; E.coli; plasmid; vector; promoter; antigen;
 PN J62115288-A.
 PD 26-MAY-1987.
 PF 14-NOV-1985; 255971.
 PR 14-NOV-1985; JP-255971.
 PA (KAGA) KAGAKU OYOBI KESSEI.
 DR WPI; 87-183377/26.
 DR N-PSDB; N70681.
 PT Recombinant DNA contg. herpes simplex virus gene - is integrated in
 PT E.coli plasmid vector, contg. incorporated promoter region for
 PT animal cells, giving antigen.
 PS Disclosure; Fig 3(1-3); 13pp; Japanese.
 CC The gene encoding HSV gB is cloned into E.coli. To increase the
 CC efficiency of expression, the gene encoding the C-terminal region
 CC of gB, which is very hydrophobic, is removed and the remaining gene
 CC is cloned into an expression vector. The vector is composed of pBR322,
 CC SV40 early promoter and the HSV gB gene.
 CC The construct is useful for the production of a good antigen for
 CC a HSV vaccine.
 SQ Sequence 903 AA;

Query Match 74.1%; Score 43; DB 1; Length 903;
 Best Local Similarity 55.6%; Pred. No. 1.51e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 209 TAKYVRNRL 217
 :|||
 QY 48 ASKYRTN 56

RESULT 8
 ID P60244 standard; Protein; 903 AA.
 AC P60244;
 DT 31-JUL-1991 (first entry)
 DE Herpes simplex virus glycoprotein gB.
 KW HSV; gB glycoprotein; vaccine.
 OS Herpes simplex virus.
 PN EP-170169-A.
 PD 05-FEB-1986.
 PF 19-JUL-1985; 109042.
 PR 11-DEC-1984; JP-282465.
 PR 20-DEC-1984; JP-151766.
 PA (KAGA) CEMO-SERO-THERAP.
 PI Nozaki C, Makizumi K, Kino Y, Eto, Ohtomo N;
 DR WPI; 86-036935/06.
 DR N-PSDB; N60195.
 PT Recombinant DNA containing herpes simplex virus gene or fragment -
 PT useful as transformant yeast for prodn. of high purity herpes
 PT simplex virus glycoproteins for use in protective vaccines.
 PS Disclosure; Fig. 6; 53pp; English.
 CC The gB glycoprotein is useful for the production of vaccines
 CC conferring protection against herpes simplex virus infections.
 SQ Sequence 903 AA;

Query Match 74.1%; Score 43; DB 1; Length 903;
Best Local Similarity 55.6%; Pred. No. 1.51e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 209 TAKYVRNRL 217
:|||||
QY 48 ASKYSRTNL 56

RESULT 9

ID P50312 standard; protein; 903 AA.
AC P50312; 1991 (first entry)
DE Herpes simplex virus 1 glycoprotein B.
KW Herpes simplex virus 1; glycoprotein B; vaccine.
OS Herpes simplex virus 1.
PN W08504587-A.
PD 24-OCT-1985.
PF 04-APR-1985; U00587.
PR 06-APR-1984; US-597784.
PR 17-JUL-1984; US-631669.
PA (CHIR-) Chiron Corp.
PI Burke RL, Pachl C, Valenzuela PDT, Urdea MS;
DR WPI: 85-276087/44.
DR N-PSDB; N50364.
PT Recombinant herpes simplex vaccine - prepd. by expression of DNA constructs in a eukaryotic host.
PS Disclosure; Table 1 page 26-30; 80pp; English.
CC Herpes simplex virus glycoprotein B or fragments may be used in a vaccine against HSV. Dosage is 10 micrograms to 2 mg/kg. The glycoprotein DNA is expressed in an eukaryotic host, esp.
CC Saccharomyces cerevisiae, CHO cells and COS cells. Suitable plasmids are pYHS15, 116, 117, 118 and 119.
SQ Sequence 903 AA;

Query Match 74.1%; Score 43; DB 1; Length 903;
Best Local Similarity 55.6%; Pred. No. 1.51e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 210 TAKYVRNRL 218
:|||||
QY 48 ASKYSRTNL 56

RESULT 10

ID P71135 standard; protein; 903 AA.
AC P71135;
DE 30-APR-1991 (first entry)
DE Herpes Simplex Virus-1 gB from pKBXX.
KW vaccine; prophylaxis; treatment; Herpes Simplex Virus-1; glycoprotein; gB.
OS Herpes simplex virus type 1 (KOS).
PN US4642333-A.
PD 10-FEB-1987.
PF 20-JUN-1984; 622496.
PR 18-SEP-1983; US-532996.
PR 20-JUN-1985; US-622496.
PA (PERS/) PERSON S.
PI Person S;
DR WPI: 87-056354/08.
DR N-PSDB; N71303.
PT Amino acid chain of glyco:protein B of HSV-1 and 2 - prepd. as recombinant and used for vaccines for herpes simplex virus types 1 and 2.
PS Example; Table 2; 16pp; English.
CC A pure non-glycosylated amino acid (AA) chain comprising a sequence corresponding to that occurring in glycoprotein B of HSV-1 or HSV-2 virus which is antigenic to HSV-1 of HSV-2, which contains no more than 750 AA residues, and which includes AA residues 135-649 inclusive is claimed. It can be used to produce vaccines for prophylaxis and treatment of HSV-1 and HSV-2.
CC See N71399 for the HSV-2 sequence.
SQ Sequence 903 AA;

Query Match 74.1%; Score 43; DB 1; Length 903;
Best Local Similarity 55.6%; Pred. No. 1.51e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 209 TAKYVRNRL 217
:|||||
QY 48 ASKYSRTNL 56

RESULT 11

ID R14665 standard; protein; 904 AA.
AC R14665;
DT 23-JAN-1992 (first entry)
DE HSVGB polypeptide.
KW Vaccine; antigen.
OS Herpes simplex virus.
PN J03218397-A.
PD 25-SEP-1991.
PR 21-JUN-1990; J61448.
PR 22-JUN-1989; JP-158238.
PR 30-NOV-1989; JP-308941.
PR 21-JUN-1990; JP-161448.
PA (TAKE) TAKEDA CHEMICAL IND KK.
DR WPI: 91-328397/45.
DR N-PSDB; Q14478.
PT HSVGB polypeptide(s) obtd. by recombinant DNA techniques - useful as vaccines against HSV and in diagnosis, can be produced cheaply and safely.
PS Claim 1; Fig 2; 24pp; Japanese.
CC The sequence was deduced from DNA and is that of the HSVGB polypeptide. The recombinant protein or esp. truncated forms of it, can be used to prepare vaccines for prophylaxis of HSV infection and for use in diagnostic kits.
CC See also R14666.
SQ Sequence 904 AA;

Query Match 74.1%; Score 43; DB 1; Length 904;
Best Local Similarity 55.6%; Pred. No. 1.51e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 210 TAKYVRNRL 218
:|||||
QY 48 ASKYSRTNL 56

RESULT 12

ID R41779 standard; protein; 904 AA.
AC R41779;
DT 29-MAR-1994 (first entry)
DE Glycoprotein B (gB2).
KW Glycoprotein; gB1; gB2; herpes simplex virus; HSV-1; HSV-2; flanking; initiation; termination; transcription; translation; regulatory sequence.
OS Herpes simplex virus 2 strain 333.
FH Key Location/Qualifiers
FT misc_difference 82 /note= "ARG encoded by GCG (sic)"
FT misc_difference 478 /note= "ASN encoded by ACC (sic)"
FT misc_difference 540 /note= "ASN encoded by ACC (sic)"
FT misc_difference 882 /note= "ARG encoded by GCC (sic)"
FT misc_difference
PN US5244792-A.
PD 14-SEP-1993.
PF 06-APR-1984; 597784.
PR 06-APR-1984; US-597784.
PR 20-OCT-1986; US-921730.
PR 20-SEP-1990; US-587179.
PA (CHIR) CHIRON CORP.
PI Burke RL, Pachl C, Valenzuela PDT;

DR WPI; 93-302641/38.
 DR N-PSDB; Q48496.
 PT DNA construct for expressing HSV glycosylated polypeptide -
 PT useful for vaccinating against HSV-1 and -2 infections
 PS Disclosure; Fig 4; 33pp; English.
 CC DNA constructs for expressing a glycosylated polypeptide in a
 CC eukaryotic host cell comprises (1) an oligonucleotide sequence (OS)
 CC free of natural flanking sequences, encoding glycoprotein B (gB)
 CC of HSV, or its fragments, and (2) 5' initiating and 3' terminating
 CC transcriptional and translational regulatory sequences flanking OS,
 CC at least one of these sequences not being from HSV. The HSV-1 and
 CC HSV-2 sequences are given in Q48496 and Q48497 respectively.
 SQ Sequence 904 AA;

Query Match 74.1%; Score 43; DB 1; Length 904;
 Best Local Similarity 55.6%; Pred. No. 1.51e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 210 TAKYVRNRL 218
 :||| |||
 QY 48 ASKYSRTNL 56

RESULT 13
 ID W34533 standard; Protein; 904 AA.
 AC W34533;
 DT 26-FEB-1998 (first entry)
 DE Herpes simplex virus type 2 glycoprotein B.
 KW Herpes simplex virus; HSV; type I; type II; glycoprotein B; vaccine;
 KW membrane protein; virus-specific glycoprotein;
 KW transmembrane anchor region.
 OS Herpes simplex virus type 2.
 PN US5648079-A.
 PD 15-JUL-1997.
 PF 08-DEC-1994; 351875.
 PR 20-OCT-1986; US-921730.
 PR 06-APR-1984; US-597784.
 PR 20-SEP-1990; US-587179.
 PR 21-DEC-1992; US-993415.
 PR 18-OCT-1993; US-138717.
 PA (CHIR) CHIRON CORP.
 PI Burke RL, Pachl C, Valenzuela PDT;
 DR WPI; 97-372022/34.
 DR N-PSDB; T93651.
 PT Vaccines against herpes simplex virus infection - containing
 PT recombinant HSV glycoprotein B
 PS Disclosure; Fig 4A-F; 33pp; English.
 CC The present sequence is from the Herpes simplex virus (HSV) type 2
 CC strain 333 and represents a glycoprotein B. HSV is a double stranded
 CC virus packaged within an icosahedral nucleocapsid enveloped within a
 CC membrane. The membrane includes a number of virus-specific
 CC glycoproteins, with glycoprotein B being one of the most abundant.
 CC Glycoprotein B from both HSV type I and type II are cross reactive. New
 CC vaccines comprising recombinantly produced glycosylated glycoprotein B
 CC that has a deletion of at least part of the transmembrane anchor region,
 CC in combination with a carrier and an adjuvant have been produced. The
 CC vaccines are used for immunising humans against HSV (HSV type 1 or
 CC HSV type 2) by vaccination before or after primary infection with HSV.
 SQ Sequence 904 AA;

Query Match 74.1%; Score 43; DB 1; Length 904;
 Best Local Similarity 55.6%; Pred. No. 1.51e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 210 TAKYVRNRL 218
 :||| |||
 QY 48 ASKYSRTNL 56

RESULT 14
 ID W00375 standard; Protein; 904 AA.
 AC W00375; 1997 (first entry)
 DT 24-JAN-1997

DE HSV-1 glycoprotein B.
 KW HSV-1; glycoprotein B; vaccine; diagnosis; immunoassay.
 OS Herpes simplex virus type 1.
 PH Key Location/Qualifiers
 FT peptide 1..22
 FT /label= Sig_peptide
 FT peptide 14..119
 FT /label= Dominant_type-specific_epitope
 FT domain 23..744
 FT /label= Extracellular_domain
 FT peptide 295..507
 FT /label= Type-specific_epitope
 FT domain 745..798
 FT /label= Transmembrane_domain
 FT domain 799..804
 FT /label= Cytoplasmic_domain
 FT peptide 814..901
 FT /label= Cross-reactive_antigenic_epitope
 PN W09632962-A1.
 PD 24-OCT-1996.
 PR 19-APR-1996; U05316.
 PR 21-APR-1995; US-426604.
 PA (UYNE-) UNIV NEW MEXICO STATE.
 PI Bell R, Goade DE, Jenison S;
 DR WPI; 96-485357/48.
 PT New type-specific and cross-reactive Herpes Simplex Virus epitope(s)
 PT - used for diagnosing HSV-1 and/or HSV-2 infections and in vaccines
 PT for producing neutralising antibodies
 PS Disclosure; Fig 4; 37pp; English.
 CC Herpes simplex virus type 1 (HSV-1) glycoprotein B (gB1) (W00375)
 CC and HSV-2 gB2 (W00376) include type-specific and cross-reactive
 CC antigenic epitopes. Recombinant type-specific epitopes can be used
 CC in immunoassays to distinguish between HSV-1 and HSV-2 infections.
 CC Recombinant cross-reactive epitopes are useful for the diagnosis of
 CC both HSV-1 and HSV-2 infections. Both types of epitope can be used
 CC in vaccines.
 SQ Sequence 904 AA;

Query Match 74.1%; Score 43; DB 1; Length 904;
 Best Local Similarity 55.6%; Pred. No. 1.51e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 210 TAKYVRNRL 218
 :||| |||
 QY 48 ASKYSRTNL 56

RESULT 15
 ID P80915 standard; Protein; 905 AA.
 AC P80915;
 DT 09-MAR-1992 (first entry)
 DE Sequence of Herpes simplex virus (HSV) glycoprotein B (gB) 2.
 KW vaccine; herpes simplex virus; therapy.
 OS Herpes simplex virus.
 PN W0802634-A.
 PD 21-APR-1988.
 PF 20-OCT-1987; U02709.
 PR 20-OCT-1986; US-921213.
 PR 20-JUL-1987; US-079605.
 PA (CHIR-) CHIRON CORP.
 PI Burke LR, Pachl C, Valenzuela P;
 DR WPI; 88-119368/17.
 DR N-PSDB; N80908.
 PT Vaccine for treatment of herpes simplex virus - contains
 PT recombinant HSV glyco:proteins B and D
 PS Example; Fig 4; 71pp; English.
 CC Prepn. of recombinant gB and gD is described in W088504587. The
 CC amino acid sequence and DNA sequence for gB 1 presented in P80914
 CC and N80907 differ from that originally presented in Table 1 of
 CC International Publication No. WO 85/04587. The DNA sequence in Table
 CC 1 contains an error in that an additional nucleotide (G) is listed
 CC at position 607 which resulted in a shift in reading frame relative
 CC to N80907 from which this nucleotide has been deleted.

SQ Sequence 905 AA;

Query Match 74.1%; Score 43; DB 1; Length 905;
 Best Local Similarity 55.6%; Pred. No. 1.51e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 209 TAKYVRNRL 217

QY 48 ASKYSRTNL 56

Search completed: Wed May 10 12:26:10 2000
 Job time : 8 secs.

Query Match 74.1%; Score 43; DB 2; Length 51;
Best Local Similarity 71.4%; Pred. No. 8.51e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 14 KYARTAL 20
11:111
QY 50 KYSRTNL 56

RESULT 2
ID PCT-US95-14442A-91 STANDARD; PRT; 255 AA.
XX
AC xxxxxx
XX
DT
XX
DE

Sequence 91, Application PC/TUS9514442A

Sequence 91, Application PC/TUS9514442A

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.

APPLICANT: Rushlow, Keith E.

APPLICANT: Hunter, Shirley Wu

APPLICANT: Frank, Glenn R.

APPLICANT: Heath, Andrew W.

APPLICANT: Yamaka, Miles Yamanaka

APPLICANT: Arfsten, Ann

APPLICANT: Dale, Beverly

APPLICANT: Stiegler, Gary

TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND

TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA

TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID

TITLE OF INVENTION: MOLECULES, AND USES THEREOF

NUMBER OF SEQUENCES: 119

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross & McIntosh

STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver

STATE: Colorado

COUNTRY: USA

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/14442A

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Gary J. Connell

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 91:

SEQUENCE CHARACTERISTICS:

LENGTH: 255 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 255 AA; 28095 MW; 355595 CN;

Query Match 74.1%; Score 43; DB 3; Length 255;
Best Local Similarity 71.4%; Pred. No. 8.51e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 105 KYDRANL 111
111111
QY 50 KYSRTNL 56

Query Match 74.1%; Score 43; DB 2; Length 51;
Best Local Similarity 71.4%; Pred. No. 8.51e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 14 KYARTAL 20
11:111
QY 50 KYSRTNL 56

RESULT 2
ID PCT-US95-14442A-91 STANDARD; PRT; 255 AA.
XX
AC xxxxxx
XX
DT
XX
DE

Sequence 91, Application PC/TUS9514442A

Sequence 91, Application PC/TUS9514442A

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.

APPLICANT: Rushlow, Keith E.

APPLICANT: Hunter, Shirley Wu

APPLICANT: Frank, Glenn R.

APPLICANT: Heath, Andrew W.

APPLICANT: Yamaka, Miles Yamanaka

APPLICANT: Arfsten, Ann

APPLICANT: Dale, Beverly

APPLICANT: Stiegler, Gary

TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND

TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA

TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID

TITLE OF INVENTION: MOLECULES, AND USES THEREOF

NUMBER OF SEQUENCES: 119

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross & McIntosh

STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver

STATE: Colorado

COUNTRY: USA

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/14442A

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Gary J. Connell

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 91:

SEQUENCE CHARACTERISTICS:

LENGTH: 255 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 255 AA; 28095 MW; 355595 CN;

RESULT 3
ID US-08-413-118-8 STANDARD; PRT; 903 AA.
XX
AC xxxxxx
XX
DT
XX
DE

Sequence 8, Application US/08413118
Sequence 8, Application US/08413118
Patent No. 5688920
GENERAL INFORMATION:
APPLICANT: PAOLETTI, ENZO
APPLICANT: LIMBACH, KEITH J.
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
TITLE OF INVENTION: CANINE HERPESVIRUS GB, GC, AND GD AND USES THEREFOR
NUMBER OF SEQUENCES: 128
CORRESPONDENCE ADDRESS:
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.
STREET: 530 FIFTH AVENUE, 25TH FLOOR
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/413,118
FILING DATE: 29-MAR-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/220,151
FILING DATE: 30-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: FROMMER, WILLIAM S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2670
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
SEQUENCE 903 AA; 100473 MW; 4090268 CN;

Query Match 74.1%; Score 43; DB 1; Length 903;
Best Local Similarity 55.6%; Pred. No. 8.51e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 209 TAKYVRNLL 217
111111
QY 48 ASKYSRTNL 56

RESULT 4
ID US-08-220-151-8 STANDARD; PRT; 903 AA.
XX
AC xxxxxx
XX
DT
XX
DE

Sequence 8, Application US/08220151

Sequence 8, Application US/08220151

GENERAL INFORMATION:

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

APPLICANT: FROMMER, WILLIAM S.

CC Sequence 8, Application US/08220151
 CC Patent No. 5529780
 CC GENERAL INFORMATION:
 CC APPLICANT: Paolotti, Enzo
 CC APPLICANT: Limbach, Keith J.
 CC TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
 CC TITLE OF INVENTION: CANINE HERPESVIRUS gB, gC AND gD AND USES THEREFOR
 CC NUMBER OF SEQUENCES: 91
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Curtis, Morris & Safford
 CC STREET: 530 Fifth Avenue
 CC CITY: New York
 CC STATE: NY
 CC COUNTRY: USA
 CC ZIP: 10036
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patent In Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/220,151
 CC FILING DATE: 30-MAR-1994
 CC CLASSIFICATION: 435
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Frommer, William S.
 CC REGISTRATION NUMBER: 25,506
 CC REFERENCE/DOCKET NUMBER: 454310-2540
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (212) 840-3333
 CC TELEFAX: (212) 840-0712
 CC TELEX: 425066 CURTMS
 CC INFORMATION FOR SEQ ID NO: 8:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 903 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 CC FRAGMENT TYPE: N-terminal
 CC SQ SEQUENCE 903 AA; 100473 MW; 4090268 CN;
 Query Match 74.1%; Score 43; DB 1; Length 903;
 Best Local Similarity 55.6%; Pred. No. 8.51e+01;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 209 TAKYVRNRL 217
 QY 48 ASKYSRTNL 56
 RESULT 5
 ID 5244792-4 STANDARD; PRT; 979 AA.
 XX xxxxxx
 DT 01-JAN-1900
 XX Patent No. 5244792.
 DE Patent No. 5244792
 CC APPLICANT: BURKE, RAE L.; PACHL, CAROL; VALENZUELA, PABLO D.T.
 CC TITLE OF INVENTION: EXPRESSION OF RECOMBINANT GLYOPROTEIN
 CC B FROM HERPES SIMPLEX VIRUS
 CC NUMBER OF SEQUENCES: 19
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/587,179
 CC FILING DATE: 20-SEP-1990
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: 921,730
 CC FILING DATE: 20-OCT-1986
 CC APPLICATION NUMBER: 597,784
 CC FILING DATE: 06-APR-1984

CC SEQ ID NO:4:
 CC LENGTH: 904
 CC SEQUENCE 979 AA; 108375 MW; 5267201 CN;
 Query Match 74.1%; Score 43; DB 4; Length 904;
 Best Local Similarity 55.6%; Pred. No. 8.51e+01;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 210 TAKYVRNRL 218
 QY 48 ASKYSRTNL 56
 RESULT 6
 ID PCT-US96-05316-1 STANDARD; PRT; 904 AA.
 XX xxxxxx
 AC xxxxxx
 DT
 XX Sequence 1, Application PC/TUS9605316
 DE DE
 XX Sequence 1, Application PC/TUS9605316
 CC GENERAL INFORMATION:
 CC APPLICANT: Goade, Diane E.
 CC APPLICANT: Bell, Richard
 CC APPLICANT: Jenison, Steven
 CC TITLE OF INVENTION: Immunoassay for Herpes Simplex Virus.
 CC NUMBER OF SEQUENCES: 2
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Hoffman, Wasson & Gitler
 CC STREET: 2361 Jefferson Davis Highway
 CC CITY: Arlington
 CC STATE: Virginia
 CC COUNTRY: USA
 CC ZIP: 22202
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Diskette, 3.50 inch, 500 Kb storage
 CC COMPUTER: Aopen Pentium
 CC OPERATING SYSTEM: WordPerfect 5.1
 CC SOFTWARE: WordPerfect
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: PCT/US96/05316
 CC FILING DATE:
 CC CLASSIFICATION:
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: 08/426,604
 CC FILING DATE: 21-APR-95
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Gitler, Stewart L.
 CC REGISTRATION NUMBER: 31,256
 CC REFERENCE/DOCKET NUMBER: A5144PCT.SQ
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (703)415-0100
 CC TELEFAX: (703)418-2768
 CC INFORMATION FOR SEQ ID NO: 1:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 904 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC PUBLICATION INFORMATION:
 CC RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 14 TO 110
 CC RELEVANT RESIDUES IN SEQ ID NO: FROM 295 TO 507
 CC RELEVANT RESIDUES IN SEQ ID NO: FROM 814 TO 901
 CC SEQUENCE 904 AA; 100292 MW; 4088146 CN;
 Query Match 74.1%; Score 43; DB 3; Length 904;
 Best Local Similarity 55.6%; Pred. No. 8.51e+01;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 210 TAKYVRNRL 218
 QY 48 ASKYSRTNL 56

CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Schwedler, Carl J.
REGISTRATION NUMBER: 36,924
REFERENCE/DOCKET NUMBER: 15593/01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 530-792-2265
TELEFAX: 530-792-2463
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 477 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
SEQUENCE 477 AA: 52754 MW; 1139230 CN;

Query Match 72.4%; Score 42; DB 2; length 477;
Best local Similarity 55.6%; Pred. NO. 1.13e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DB 262 ASKYTRDTL 270
11:1:1 |
QY 48 ASKYSTRNL 56

RESULT 11
ID US-08-469-202-12 STANDARD; PRT; 477 AA.
XX AC xxxxxx
XX DT XX
XX DT XX
DE DE
XX XX
Sequence 12, Application US/08469202
Sequence 12, Application US/08469202
Patent No. 5750875
GENERAL INFORMATION:
APPLICANT: STALKER, DAVID
TITLE OF INVENTION: GLYCOGEN BIOSYNTHETIC ENZYMES IN PLANTS
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: CALGENE, INC.
STREET: 1920 FIFTH STREET
CITY: DAVIS
STATE: CA
COUNTRY: USA
ZIP: 95616
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.1
SOFTWARE: Microsoft Word 5.1 (a)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,202
FILING DATE: 6-JUNE-95
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/016,881
FILING DATE: 11-FEB-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Elizabeth Lassen
REGISTRATION NUMBER: 31,845
NAME: Donna E. Scherer
REGISTRATION NUMBER: 34,719
NAME: Carl J. Schwedler
REGISTRATION NUMBER: 36,924
REFERENCE/DOCKET NUMBER: CGNE 93-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 916-753-1510
TELEFAX: 916-753-1510
INFORMATION FOR SEQ ID NO: 12:

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 477 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
SQ SEQUENCE 477 AA: 52754 MW: 1139230 CN;

Query Match 72.4%; Score 42; DB 1; Length 477;
Best Local Similarity 55.6%; Pred. No. 1.13e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 262 ASRYTRDTL 270
||:|:| |
QY 48 ASKYSRTNL 56

RESULT 12
ID US-08-190-802A-68 STANDARD; PRT; 798 AA.

XX
AC xxxxxx

DE
XX Sequence 68, Application US/08190802A
XX Patent No. 5519003

CC GENERAL INFORMATION:
CC APPLICANT: Mochly-Rosen, Daria
CC APPLICANT: Ron, Dorit
CC TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
CC TITLE OF INVENTION: Thereof
CC NUMBER OF SEQUENCES: 265
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Dehlinger & Associates
CC STREET: P.O. Box 60850
CC CITY: Palo Alto
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94306-0850

CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/190,802A
CC FILING DATE: 01-FEB-1994
CC CLASSIFICATION: 530

CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fabian, Gary R.
CC REGISTRATION NUMBER: 33,875
CC REFERENCE/DOCKET NUMBER: 8600-0139
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 324-0880
CC TELEFAX: (415) 324-0960
CC INFORMATION FOR SEQ ID NO: 68:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 798 amino acids
CC TYPE: amino acid
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC ORIGINAL SOURCE:
CC INDIVIDUAL ISOLATE: yrb 1410 yeast, Fig. 51
CC SEQUENCE 798 AA: 88967 MW: 3338942 CN;

Query Match 72.4%; Score 42; DB 1; Length 798;
Best Local Similarity 71.4%; Pred. No. 1.13e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 783 KFSRSNL 789
|:|:| |
QY 50 KYSRTNL 56

RESULT 13
ID US-08-190-802A-64 STANDARD; PRT; 798 AA.

XX
AC xxxxxx

DE
XX Sequence 64, Application US/08190802A
XX Patent No. 5519003

CC GENERAL INFORMATION:
CC APPLICANT: Mochly-Rosen, Daria
CC APPLICANT: Ron, Dorit
CC TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
CC TITLE OF INVENTION: Thereof
CC NUMBER OF SEQUENCES: 265
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Dehlinger & Associates
CC STREET: P.O. Box 60850
CC CITY: Palo Alto
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94306-0850

CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/190,802A
CC FILING DATE: 01-FEB-1994
CC CLASSIFICATION: 530

CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fabian, Gary R.
CC REGISTRATION NUMBER: 33,875
CC REFERENCE/DOCKET NUMBER: 8600-0139
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 324-0880
CC TELEFAX: (415) 324-0960
CC INFORMATION FOR SEQ ID NO: 64:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 798 amino acids
CC TYPE: amino acid
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC ORIGINAL SOURCE:
CC INDIVIDUAL ISOLATE: TUP1 HOMOLOG, Fig. 47
CC SEQUENCE 798 AA: 88967 MW: 3338942 CN;

Query Match 72.4%; Score 42; DB 1; Length 798;
Best Local Similarity 71.4%; Pred. No. 1.13e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 783 KFSRSNL 789
|:|:| |
QY 50 KYSRTNL 56

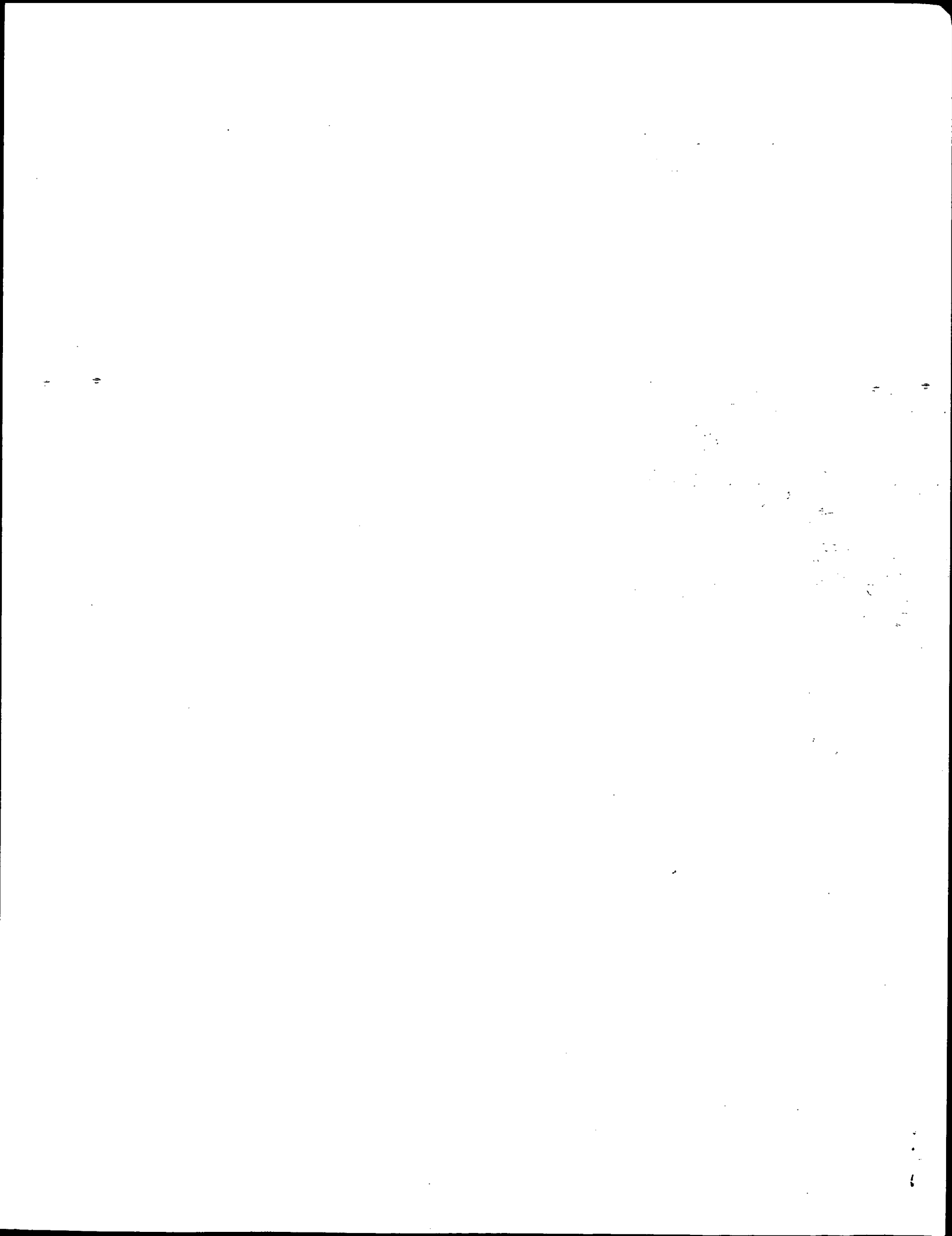
RESULT 14
ID US-08-308-818-2 STANDARD; PRT; 798 AA.

XX
AC xxxxxx

DE
XX
XX
DT
XX

DE Sequence 2, Application US/08308818
 CC Sequence 2, Application US/08308818
 CC Patent No. 5847077
 CC GENERAL INFORMATION:
 CC APPLICANT: Green, Michael R
 CC APPLICANT: Reese, Joseph C
 CC TITLE OF INVENTION: A No. 5847077el Fungal Multisubunit Protein
 CC TITLE OF INVENTION: Complex Critical for Expression of Fungal Proteins
 CC NUMBER OF SEQUENCES: 11
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Darby & Darby PC
 CC STREET: 805 Third Avenue
 CC CITY: New York
 CC STATE: New York
 CC COUNTRY: US
 CC ZIP: 10022
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/308,818
 CC FILING DATE: 19-SEP-1994
 CC CLASSIFICATION: 435
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Ludwig, S. Peter
 CC REGISTRATION NUMBER: 25,351
 CC REFERENCE/DOCKET NUMBER: 0342/0A404
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 212-52707700
 CC TELEFAX: 212-753-6237
 CC TELEX: 236687
 CC INFORMATION FOR SEQ ID NO: 2:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 798 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC HYPOTHETICAL: NO
 CC FRAGMENT TYPE: N-terminal
 CC ORIGINAL SOURCE:
 CC ORGANISM: S. cerevisiae
 CC IMMEDIATE SOURCE:
 CC CLONE: TAF-90
 CC SEQUENCE 798 AA; 88967 MW; 3338942 CN;
 Query Match 72.4%; Score 42; DB 2; Length 798;
 Best Local Similarity 71.4%; Pred. No. 1.13e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 783 KFSRNL 789
 QY 50 KYSRNL 56
 RESULT 15
 ID US-08-418-893D-1 STANDARD; PRT; 51 AA.
 AC xxxxxx
 XX
 DT
 XX
 DE Sequence 1, Application US/08418893D
 CC Sequence 1, Application US/08418893D
 CC Patent No. 5559220
 CC GENERAL INFORMATION:
 CC APPLICANT: ROESSLER, PAUL G
 CC APPLICANT: OHLROGGE, JOHN B
 CC TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
 CC TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTHELLA CRYPTICA

CC NUMBER OF SEQUENCES: 25
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
 CC STREET: 1617 Cole Blvd.
 CC CITY: Golden
 CC STATE: CO
 CC COUNTRY: USA
 CC ZIP: 80401-3393
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/418,893D
 CC FILING DATE: April 7, 1995
 CC CLASSIFICATION: 800
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 08/104,938
 CC FILING DATE: September 14, 1993
 CC CLASSIFICATION: 800
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: O'CONNOR, EDNA
 CC REGISTRATION NUMBER: 29,252
 CC REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 303-231-1000
 CC TELEFAX: 303-231-1098
 CC TELEX:
 CC INFORMATION FOR SEQ ID NO: 1:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 51 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC HYPOTHETICAL: NO
 CC ANTI-SENSE: NO
 CC FRAGMENT TYPE: internal
 CC SEQUENCE 51 AA; 5647 MW; 12698 CN;
 Query Match 70.7%; Score 41; DB 1; Length 51;
 Best Local Similarity 75.0%; Pred. No. 1.50e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 29 ASKYAREN 36
 QY 48 ASKYRSTN 55
 Search completed: Wed May 10 12:27:20 2000
 Job time: 53 secs.



US-09-376-430-2-10.ipi

Thu May 11 06:49:47 2000

 M P E R L H

 (TM)

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

protein - protein database search, using Smith-Waterman algorithm

MPSrch_PP protein - protein database search, using Smith-Waterman algorithm
 Wed May 10 12:25:37 2000: MasPar time 3.97 Seconds
 107.008 Million cell updates/sec

Run On:

Tabular output not generated.

Title: >US-09-376-430-2
 (48-56) from US09376430A.pep (10 of 25)

Description: 58

Perfect Score: 1 ASKYSTRL 9

Sequence: PAM 150

Scoring table: Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: pir62

1:pir1 2:pir2 3:pir3 4:pir4

Mean 22.169: Variance 23.402: scale 0.947

Statistics:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Pred. No.

Result No. Score Match Length DB ID Description

1 50 86.2 946 2 S64745 SF11 protein - yeast

2 48 82.8 575 2 S12525 pTD protein - Erwinia

3 47 81.0 671 1 C69621 fructose-bisphosphatase

4 46 79.3 230 2 D72070 outer membrane protein

5 46 79.3 230 2 D72070 ubiquinone methyltran

6 46 79.3 230 2 D72070 2-acetylglucosaminidase

7 46 79.3 230 2 D72070 2-acetylglucosaminidase

8 46 79.3 230 2 D72070 2-acetylglucosaminidase

9 46 79.3 230 2 D72070 2-acetylglucosaminidase

10 46 79.3 230 2 D72070 2-acetylglucosaminidase

11 46 79.3 230 2 D72070 2-acetylglucosaminidase

12 46 79.3 230 2 D72070 2-acetylglucosaminidase

13 46 79.3 230 2 D72070 2-acetylglucosaminidase

14 46 79.3 230 2 D72070 2-acetylglucosaminidase

15 46 79.3 230 2 D72070 2-acetylglucosaminidase

16 46 79.3 230 2 D72070 2-acetylglucosaminidase

17 46 79.3 230 2 D72070 2-acetylglucosaminidase

18 46 79.3 230 2 D72070 2-acetylglucosaminidase

19 46 79.3 230 2 D72070 2-acetylglucosaminidase

20 46 79.3 230 2 D72070 2-acetylglucosaminidase

21 46 79.3 230 2 D72070 2-acetylglucosaminidase

22 46 79.3 230 2 D72070 2-acetylglucosaminidase

23 46 79.3 230 2 D72070 2-acetylglucosaminidase

24 46 79.3 230 2 D72070 2-acetylglucosaminidase

25 46 79.3 230 2 D72070 2-acetylglucosaminidase

26 46 79.3 230 2 D72070 2-acetylglucosaminidase

27 46 79.3 230 2 D72070 2-acetylglucosaminidase

28 46 79.3 230 2 D72070 2-acetylglucosaminidase

29 46 79.3 230 2 D72070 2-acetylglucosaminidase

30 46 79.3 230 2 D72070 2-acetylglucosaminidase

31 46 79.3 230 2 D72070 2-acetylglucosaminidase

32 46 79.3 230 2 D72070 2-acetylglucosaminidase

33 46 79.3 230 2 D72070 2-acetylglucosaminidase

34 46 79.3 230 2 D72070 2-acetylglucosaminidase

35 46 79.3 230 2 D72070 2-acetylglucosaminidase

36 46 79.3 230 2 D72070 2-acetylglucosaminidase

37 46 79.3 230 2 D72070 2-acetylglucosaminidase

38 46 79.3 230 2 D72070 2-acetylglucosaminidase

39 46 79.3 230 2 D72070 2-acetylglucosaminidase

40 46 79.3 230 2 D72070 2-acetylglucosaminidase

41 46 79.3 230 2 D72070 2-acetylglucosaminidase

42 46 79.3 230 2 D72070 2-acetylglucosaminidase

43 46 79.3 230 2 D72070 2-acetylglucosaminidase

44 46 79.3 230 2 D72070 2-acetylglucosaminidase

45 46 79.3 230 2 D72070 2-acetylglucosaminidase

RNA polymerase II tra 3.56e+01

neurofilament triplet 3.56e+01

glycoprotein B precu 3.56e+01

glycoprotein B precu 3.56e+01

glycoprotein B precu 3.56e+01

glycoprotein B precu 3.56e+01

death-associated prot 3.56e+01

glucosyltransferase (3.56e+01

supervillin P205 - bo 3.56e+01

hypothetical protein 5.85e+01

hypothetical protein 5.85e+01

hypothetical protein 5.85e+01

E2 protein - human pa 5.85e+01

hypothetical protein 5.85e+01

transforming protein 5.85e+01

ADP-glucose--starchba 5.85e+01

TATA box-binding prot 5.85e+01

gene coI intron 1 pro 5.85e+01

tail-host specificity 5.85e+01

orf6, Lactobacillus 9.51e+01

lin-44 protein precu 9.51e+01

GMP synthase (glutami 9.51e+01

RESULTS

ENTRY

642 2 S31285

74.1 642 2 S31285

858 858

74.1 858

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

903 903

74.1 903

Best Local Similarity 88.9%; Pred. No. 8.58e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 476 ASDYSRTNL 484
QY 48 ASKYSRTNL 56

RESULT 2 S12525 #type complete
ENTRY ptD protein - Erwinia chrysanthemi
ORGANISM #formal_name Erwinia chrysanthemi
DATE 21-Nov-1993 #sequence_revision 21-Jul-1995 #text_change
ACCESSIONS S12525
REFERENCE S12525
#authors Letoffe, S.; Delepeleire, P.; Wandersman, C.
#journal BMO J. (1990) 9:1375-1382
#title Protease secretion by Erwinia chrysanthemi: the specific
coli alpha-haemolysin.
#cross-references MUID:9028333
#accession S12525
#status Preliminary
#molecule_type DNA
#residues 1-575 #label LET
#cross-references #superfamily unassigned ATP-binding cassette proteins;
CLASSIFICATION #ATP; P-loop; transmembrane protein
#domain ATP-binding cassette homology #label ABC\
#region 575 nucleotide-binding motif A.(P-loop)
#length 575 #molecular-weight 61617 #checksum 5527
KEYWORDS 82.8%; Score 48; DB 2; Length 575;
Best Local Similarity 77.8%; Pred. No. 2.59e+00;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 244 ASKYSRTNL 252
QY 48 ASKYSRTNL 56

RESULT 3 C69621 #type complete
ENTRY fructose-bisphosphatase [EC 3.1.3.11] - Bacillus subtilis
TITLE #formal_name Bacillus subtilis
ORGANISM 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
DATE C69621
ACCESSIONS A69580
REFERENCE Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.; Entian, K.D.; Ertter, J.;
Fabret, C.; Ferrari, E.; Fuma, S.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Gasser, P.; Goffeau, A.; Gallon, N.; Ghim,
S.Y.; Glaser, P.; Guy, B.J.; Haga, K.; Haech, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Haech, E.J.; Grandi, G.;
C.P.; Henaut, A.; Hilbert, H.; Holsappel, J.; Harwood, G.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, S.; Hosono, S.;
Kasahara, Y.; Klagsbrunn, J.; Klein, C.; Karamata, D.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Krogh, S.; Kumano, M.;
Lazarevic, V.; Lee, C.M.; Levine, A.; Liu, H.; Masuda, S.;
Maeel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,

V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivoita, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yoshida, K.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession C69621
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-671 #label KUN
#cross-references GB:293124; GB:AL009126; NID:92636442;
#experimental_source strain 168
GENETICS
#gene fbp; yvde
CLASSIFICATION #superfamily Bacillus subtilis fructose-bisphosphatase;
KEYWORDS phosphoenolpyruvate carboxylase
glucosyltransferase; metallopeptidase; phosphoric monoester
SUMMARY #length 671 #molecular-weight 77978 #checksum 2936
Query Match 81.0%; Score 47; DB 1; Length 671;
Best Local Similarity 55.6%; Pred. No. 4.45e+00;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 154 SSKYTRSKL 162
QY 48 ASKYSRTNL 56

RESULT 4 S69781 #type complete
ENTRY outer membrane protein 1 variant E precursor - Dichelobacter
TITLE #formal_name Dichelobacter nodosus
ORGANISM 29-Jul-1997 #sequence_revision 29-Jul-1997 #text_change
DATE S69781
ACCESSIONS S69780
REFERENCE Moses, E.K.; Good, R.T.; Sinistaj, M.; Billington, S.J.;
Langford, C.J.; Rood, J.I.
Mol. Microbiol. (1995) 17:183-196
A multiple site-specific DNA-inversion model for the control
of Omp1 phase and antigenic variation in Dichelobacter
#cross-references MUID:96020672
#accession S69781
#status nucleic acid sequence not shown
#molecule_type DNA
#residues 1-707 #label MOS
#cross-references EMBL:U02462
GENETICS
#gene omp1
FEATURE
1-18
19-707
SUMMARY
#domain signal sequence #status predicted #label SIG
#product outer membrane protein 1 variant E #status
#length 707 #molecular-weight 76267 #checksum 1694
Query Match 81.0%; Score 47; DB 2; Length 707;
Best Local Similarity 75.0%; Pred. No. 4.45e+00;

RESULT 5 S69781 #type complete
ENTRY outer membrane protein 1 variant E precursor - Dichelobacter
TITLE #formal_name Dichelobacter nodosus
ORGANISM 29-Jul-1997 #sequence_revision 29-Jul-1997 #text_change
DATE S69781
ACCESSIONS S69780
REFERENCE Moses, E.K.; Good, R.T.; Sinistaj, M.; Billington, S.J.;
Langford, C.J.; Rood, J.I.
Mol. Microbiol. (1995) 17:183-196
A multiple site-specific DNA-inversion model for the control
of Omp1 phase and antigenic variation in Dichelobacter
#cross-references MUID:96020672
#accession S69781
#status nucleic acid sequence not shown
#molecule_type DNA
#residues 1-707 #label MOS
#cross-references EMBL:U02462
GENETICS
#gene omp1
FEATURE
1-18
19-707
SUMMARY
#domain signal sequence #status predicted #label SIG
#product outer membrane protein 1 variant E #status
#length 707 #molecular-weight 76267 #checksum 1694
Query Match 81.0%; Score 47; DB 2; Length 707;
Best Local Similarity 75.0%; Pred. No. 4.45e+00;

US-09-376-430-2-10.rpt

Thu May 11 06:49:47 2000

```

Matches      6; Conservative      1; Mismatches      1; Indels      0; Gaps      0;

Db 298 AKYERTNL 305
    :||| |||
QY 49 SKYSRTNL 56

RESULT 5
ENTRY #type complete
TITLE ubiquitinone methyltransferase - Chlamydia pneumoniae (strain CWL029)
ORGANISM #formal_name Chlamydia pneumoniae
DATE 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Sep-1999
ACCESSIONS A72070
REFERENCE Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Davis, R.W.; Stephens, R.S.
#authors Nature Genet. (1999) 21:385-389
#journal Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
#title #cross-references MUID:99206606
#accession A72070
#status preliminary
#molecule_type DNA
#residues 1-230 #label ARN
#cross-references GB:AE001636; GB:AE001363; NID:g4376794; PID:g4376803
#experimental_source strain CWL029
GENETICS ubiE
#gene #superfamily spore germination protein C2; bioc homology
CLASSIFICATION #length 230 #molecular-weight 25637 #checksum 4715
SUMMARY
Query Match 79.3%; Score 46; DB 2; Length 230;
Best Local Similarity 87.5%; Pred. No. 7.59e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 18 ASKYDRTN 25
    :||| |||
QY 48 ASKYSRTN 55

RESULT 6
ENTRY #type complete
TITLE hypothetical protein TM1021 - Thermotoga maritima (strain MSB8)
ORGANISM #formal_name Thermotoga maritima
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
ACCESSIONS D72302
REFERENCE Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek, J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
#journal Nature (1999) 399:323-329
#title Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.
#cross-references MUID:99287316
#accession D72302
#status preliminary
#molecule_type DNA
#residues 1-422 #label ARN
#cross-references GB:AE001764; GB:AE000512; NID:g4981561; PID:g4981562; TIGR:TM1021
#experimental_source strain MSB8
GENETICS
#gene
SUMMARY
TM1021
#length 422 #molecular-weight 46050 #checksum 5193

Query Match 79.3%; Score 46; DB 2; Length 422;
Best Local Similarity 66.7%; Pred. No. 7.59e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 60 ADKYSRKNL 68
    :||| |||
QY 48 ASKYSRTNL 56

RESULT 7
ENTRY #type complete
TITLE hypothetical protein CF632 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
ORGANISM #formal_name Chlamydia trachomatis
DATE 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 21-Nov-1998
ACCESSIONS D71489
REFERENCE Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov, R.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
#authors Science (1998) 282:754-759
#journal Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.
#title #cross-references MUID:99000809
#accession D71489
#status preliminary
#molecule_type DNA
#residues 1-529 #label ARN
#cross-references GB:AE001334; GB:AE001273; NID:g3329078; PID:g3329080
#experimental_source serotype D, strain UW-3/Cx
GENETICS CT632
#gene #molecular-weight 60915 #checksum 6502
SUMMARY
Query Match 79.3%; Score 46; DB 2; Length 529;
Best Local Similarity 75.0%; Pred. No. 7.59e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 43 SKYSRSTL 50
    :||| |||
QY 49 SKYSRTNL 56

RESULT 8
ENTRY #type complete
TITLE 2-acylglycerophosphoethanolamine acyltransferase (aas) RP620
ORGANISM #formal_name Rickettsia prowazekii
DATE 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 07-May-1999
ACCESSIONS E71667
REFERENCE Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.; Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland, C.G.
#journal Nature (1998) 396:133-140
#title The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
#cross-references MUID:99039499
#accession E71667
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-1152 #label AND
#cross-references GB:AJ235272; GB:AJ235269; NID:g3861033; PID:el342907; PID:g3861163
#experimental_source strain Madrid E
GENETICS
#gene
SUMMARY
aas; RP620
#length 1152 #molecular-weight 130356 #checksum 5095

```

```

Query Match      79.3%; Score 46; DB 2; Length 1152;
Best Local Similarity 66.7%; Pred. No. 7.59e+00;
Matches          6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 77 ADKIERANL 85
QY 48 ASKYSRTNL 56

RESULT 9
ENTRY T01962 #type complete
TITLE glutamate decarboxylase (EC 4.1.1.15) 2, calmodulin-binding -
ORGANISM common tobacco
DATE #formal_name Nicotiana tabacum #common_name common tobacco
26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change
ACCESSIONS T01962
REFERENCE 214473
#authors Yun, S.J.; Oh, S.H.
#journal Mol. Cells (1998) 8:125-129
#title Cloning and characterization of a tobacco cDNA encoding
#cross-references MUID:98302498 calcium/calmodulin-dependent glutamate decarboxylase.
#accession T01962
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-496 #label YUN
#cross-references EMBL:AF020424; NID:g3252853; PIDN:AAC39483.1;
PID:g3252854

GENETICS
#gene GAD2
#description catalyzes the formation of gamma-aminobutyric acid, from
L-glutamic acid
CLASSIFICATION #superfamily Escherichia coli glutamate decarboxylase
KEYWORDS calmodulin binding; carbon-carbon lyase; carboxy-lyase
SUMMARY #length 496 #molecular-weight 55931 #checksum 6623

Query Match      77.6%; Score 45; DB 2; Length 496;
Best Local Similarity 66.7%; Pred. No. 1.28e+01;
Matches          6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 19 ASRYVRTSL 27
QY 48 ASKYSRTNL 56

RESULT 10
ENTRY A48767 #type complete
TITLE glutamate decarboxylase (EC 4.1.1.15), calmodulin-binding -
ORGANISM garden petunia
DATE #formal_name Petunia x hybrida #common_name garden petunia
19-May-1995 #sequence_revision 19-May-1995 #text_change
ACCESSIONS A48767
REFERENCE A48767
#authors Baum, G.; Chen, Y.; Arazzi, T.; Takatsujit, H.; Fromm, H.
#journal J. Biol. Chem. (1993) 268:19610-19617
#title A plant glutamate decarboxylase containing a calmodulin
#cross-references MUID:93374956 binding domain. Cloning, sequence, and functional analysis.
#accession A48767
#status preliminary
#molecule_type mRNA
#residues 1-500 #label BAU
#cross-references GB:L16797; NID:g294111; PIDN:AAA33709.1; PID:g294112

CLASSIFICATION #superfamily Escherichia coli glutamate decarboxylase
KEYWORDS calmodulin binding; carbon-carbon lyase; carboxy-lyase
SUMMARY #length 500 #molecular-weight 56726 #checksum 915

Query Match      77.6%; Score 45; DB 2; Length 500;
Best Local Similarity 66.7%; Pred. No. 1.28e+01;
Matches          6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 19 ASRYVRTSL 27
QY 48 ASKYSRTNL 56

RESULT 11
ENTRY H48563 #type fragment
TITLE G1 protein - fowlpox virus (strain HP444) (fragment)
ORGANISM #formal_name fowlpox virus
DATE 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change
ACCESSIONS H48563
REFERENCE A48563
#authors Binns, M.M.; Boursnell, M.E.; Skinner, M.A.
#journal Virus Res. (1992) 24:161-172
#title Gene translocations in poxviruses: the fowlpox virus
#cross-references MUID:92410746 and occupies the locus which in vaccinia virus is occupied
#accession H48563 by the ribonucleotide reductase large subunit gene.
#molecule_type DNA
#residues 1-341 #label BIN
#cross-references GB:AJ223385; NID:g3123522; PID:el292200; PID:g3123537
#note sequence extracted from NCBI backbone (NCBIN:113549,
NCBIP:113557)

GENETICS
#gene G1
#summary #length 341 #checksum 9456

Query Match      75.9%; Score 44; DB 2; Length 341;
Best Local Similarity 75.0%; Pred. No. 2.15e+01;
Matches          6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 145 SKYEKTNL 152
QY 49 SKYSRTNL 56

RESULT 12
ENTRY T16204 #type complete
TITLE hypothetical protein F28F5.3 - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
ACCESSIONS T16204
REFERENCE T16204
#authors Pauley, A.
#submission submitted to the EMBL Data Library, April 1994
#description The sequence of C. elegans cosmid F28F5.
#accession T16204
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-899 #label PAU
#cross-references EMBL:U00045; NID:g470353; PID:g470357;
PIDN:AAA50684.1; CESP:F28F5.3
#experimental_source strain Bristol N2

GENETICS
#gene CESP:F28F5.3
#introns 22/3; 41/3; 54/3; 72/2; 129/3; 182/3; 307/3; 365/1; 444/3;
525/3; 632/3; 674/1; 696/3; 784/2; 841/3; 870/3
SUMMARY #length 899 #molecular-weight 102332 #checksum 6210

Query Match      75.9%; Score 44; DB 2; Length 899;
Best Local Similarity 55.6%; Pred. No. 2.15e+01;
Matches          5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 514 ATKYPTDF 522
QY 48 ASKYSRTNL 56

```

US-09-376-430-2-10.rpr

Thu May 11 06:49:47 2000

RESULT 15

ENTRY #type complete
TITLE hypothetical protein PH1213 - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998

ACCESSIONS G71064
REFERENCE A71000

#authors Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, I.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.

#journal DNA Res. (1998) 5:55-76
#title Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.

#cross-references MUID:98344137
#accession G71064
#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA
#residues 223
#cross-references GB:AP000005; NID:g3236132; PID:d1031256; PID:g3257630
#experimental_source strain OT3
#note this accession replaces an interim accession for a sequence replaced by GenBank

GENETICS PH1213
#gene #length 223 #molecular-weight 25680 #checksum 1597
SUMMARY

Query Match 74.1%; Score 43; DB 2; Length 223;
Best Local Similarity 55.6%; Pred. No. 3.56e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 169 ACTYGRDRL 177
|:|:|:|:
QY 48 ASKYSRTNL 56

Search completed: Wed May 10 12:25:45 2000
Job time : 8 secs.

RESULT 13

ENTRY #type complete
TITLE hypothetical protein PAB3022 - Pyrococcus abyssii (strain Orsay)
ORGANISM #formal_name Pyrococcus abyssii
DATE 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999

ACCESSIONS B75193
REFERENCE A75001

#authors anonymous, Genoscope
#submission submitted to the EMBL Data Library, July 1999
#description Pyrococcus abyssii genome sequence: insights into archaeal chromosome structure and evolution.

#accession B75193
#status preliminary
#molecule_type DNA
#residues 1-78 #label KAW
#cross-references GB:AJ248283; GB:AL096836; NID:g5457433; PID:e1514887; PID:g5457502
#experimental_source strain Orsay

GENETICS PAB3022
#gene #length 78 #molecular-weight 8816 #checksum 956
SUMMARY

Query Match 74.1%; Score 43; DB 2; Length 78;
Best Local Similarity 62.5%; Pred. No. 3.56e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 71 AKYSRRDL 78
|:|:|:|:
QY 49 SKYSRTNL 56

RESULT 14

ENTRY #type complete
TITLE photosystem I chain III - Odontella sinensis chloroplast
ALTERNATE_NAMES photosystem I plastocyanin-binding chain
ORGANISM #formal_name chloroplast Odontella sinensis
DATE 17-Feb-1998 #sequence_revision 26-Feb-1998 #text_change 26-Aug-1999

ACCESSIONS S78329
REFERENCE S78238

#authors Kowalik, K.V.; Stoebe, B.; Schaffran, I.; Kroth-Pancic, P.; Freier, U.

#journal Plant Mol. Biol. Rep. (1995) 13:336-342
#title The Chloroplast Genome of a Chlorophyll a+c- containing Alga, Odontella sinensis.

#accession S78329
#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA
#residues 1-185 #label KOW
#cross-references EMBL:267753; NID:g1185127; PIDN:CAA91702.1; PID:e212166; PID:g1185219
#note the nucleotide sequence was submitted to the EMBL Data Library, November 1995

GENETICS psaf
#gene chloroplast
#genome #superfamily photosystem I chain III
CLASSIFICATION chloroplast; membrane-associated complex; photosynthesis;
KEYWORDS photosystem I
#length 185 #molecular-weight 20747 #checksum 3887

Query Match 74.1%; Score 43; DB 2; Length 185;
Best Local Similarity 71.4%; Pred. No. 3.56e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 79 KYRSSEL 85
|:|:|:|:
QY 50 KYRSRTNL 56

M P S R L H
(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 12:19:22 2000; Maspar time 90.04 Seconds
3.044 Million cell updates/sec
Tubular output not generated.

Title: >US-09-376-430-2
Description: (48-56) from US09376430A.pep (10 of 25)
Perfect Score: 58
Sequence: 1 ASKYSRTNL 9

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 22.691; Variance 21.598; scale 1.051
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description	Pred. No.
1	48	82.8	575	1	PRTD_ERWCH	PROTEASES SECRETION AT	9.30e+01
2	45	77.6	500	1	DCEL_PTHY	GLUTAMATE DECARBOXYLAS	5.24e+00
3	45	77.6	500	1	DCEL_ARATH	GLUTAMATE DECARBOXYLAS	5.24e+00
4	45	77.6	536	1	CATA_DEIRA	CATALASE (EC 1.11.1.6)	5.24e+00
5	44	75.9	789	1	RIR1_HSV4	RIBONUCLEOSIDE-DIPHOSP	9.13e+00
6	44	75.9	1551	1	YQ12_CAEEL	HYPOTHETICAL 175.7 KD	9.13e+00
7	43	74.1	185	1	PSAF_ODOSI	PHOTOSYSTEM I REACTION	1.58e+01
8	43	74.1	358	1	VE2_HPV58	REGULATORY PROTEIN E2.	1.58e+01
9	43	74.1	513	1	KA88_SCHPO	PROBABLE SERINE/THREON	1.58e+01
10	43	74.1	642	1	TFBI_YEAST	RNA POLYMERASE II TRAN	1.58e+01
11	43	74.1	857	1	NFM_CHICK	NEUROFILAMENT TRIPLET	1.58e+01
12	43	74.1	903	1	VGLB_HSV1F	GLYCOPROTEIN B PRECURS	1.58e+01
13	43	74.1	904	1	VGLB_HSV1P	GLYCOPROTEIN B PRECURS	1.58e+01
14	43	74.1	904	1	VGLB_HSV1K	GLYCOPROTEIN B PRECURS	1.58e+01
15	43	74.1	904	1	VGLB_HSV1L	GLYCOPROTEIN B-1 PRECU	1.58e+01
16	43	74.1	917	1	VGLB_HSV1H	GLYCOPROTEIN B-1 PRECU	1.58e+01
17	43	74.1	1431	1	DAPK_HUMAN	DEATH-ASSOCIATED PROTE	1.58e+01
18	42	72.4	173	1	YKNI_YEAST	HYPOTHETICAL 19.8 KD P	2.69e+01
19	42	72.4	328	1	ZIPA_HAEIN	CELL DIVISION PROTEIN	2.69e+01
20	42	72.4	353	1	VE2_HPV33	REGULATORY PROTEIN E2.	2.69e+01
21	42	72.4	477	1	GLGA_ECOLI	GLYCOGEN SYNTHASE (EC	2.69e+01
22	42	72.4	503	1	VE2_HPV21	REGULATORY PROTEIN E2.	2.69e+01
23	42	72.4	798	1	TZD4_YEAST	TRANSCRIPTION INITIATI	2.69e+01

24	41	70.7	85	1	YG36_BPMV4	HYPOTHETICAL 9.5 KD PR	4.53e+01
25	41	70.7	175	1	YMF3_CAEEL	HYPOTHETICAL 20.2 KD P	4.53e+01
26	41	70.7	231	1	Y867_RICPR	HYPOTHETICAL PROTEIN R	4.53e+01
27	41	70.7	269	1	AGAR_ECOLI	PUTATIVE AGA OPERON TR	4.53e+01
28	41	70.7	323	1	CRO_RANTE	RHO CRYSTALLIN.	4.53e+01
29	41	70.7	323	1	CRO_RANCA	RHO CRYSTALLIN.	4.53e+01
30	41	70.7	462	1	TUBE_DROME	TUBE PROTEIN.	4.53e+01
31	41	70.7	494	1	DCE2_ARATH	GLUTAMATE DECARBOXYLAS	4.53e+01
32	41	70.7	570	1	YGR0_YEAST	HYPOTHETICAL 65.8 KD P	4.53e+01
33	41	70.7	638	1	GURAA_MYCLE	GMP SYNTHASE [GLUTAMIN	4.53e+01
34	41	70.7	683	1	MAOC_ORYSA	NADP-DEPENDENT MALIC E	4.53e+01
35	41	70.7	683	1	RPC_BPPHC	REPRESSOR PROTEIN C.	4.53e+01
36	41	70.7	928	1	HXA2_HAEIN	HEME-HEMOPEXIN-BINDING	4.53e+01
37	41	70.7	1193	1	RPB2_CAEEL	DNA-DIRECTED RNA POLYM	7.56e+01
38	40	69.0	111	1	PT00_YEAST	PET100 PROTEIN PRECURS	7.56e+01
39	40	69.0	385	1	YWM_BACSU	HYPOTHETICAL 42.8 KD P	7.56e+01
40	40	69.0	416	1	PXNL_XENLA	PENTRAXIN FUSION PROTE	7.56e+01
41	40	69.0	535	1	PYRE_TREDE	PYRBI PROTEIN [INCLUDE	7.56e+01
42	40	69.0	618	1	ILVB_MYCTU	ACETOLACTATE SYNTHASE	7.56e+01
43	40	69.0	719	1	YMP8_CAEEL	HYPOTHETICAL 82.6 KD P	7.56e+01
44	40	69.0	810	1	YLZ5_CAEEL	HYPOTHETICAL 89.2 KD P	7.56e+01
45	40	69.0	928	1	YG23_YEAST	HYPOTHETICAL 106.7 KD	7.56e+01

ALIGNMENTS

RESULT ID	PRTD_ERWCH	STANDARD	PRT	575 AA.
AC	P23596			
DT	01-NOV-1991	(Rel. 20, Created)		
DT	01-NOV-1991	(Rel. 20, Last sequence update)		
DT	01-FEB-1996	(Rel. 33, Last annotation update)		
DE	PROTEASES SECRETION ATP-BINDING PROTEIN PRTD.			
GN	PRTD.			
OS	Erwinia chrysanthemi.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Erwinia..			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 9028333.			
RA	Lefloffe S., Deleplaire P., Wandersman C.;			
RA	"Protease secretion by Erwinia chrysanthemi: the specific secretion			
RT	functions are analogous to those of Escherichia coli			
RT	alpha-haemolysin."			
RL	EMBO J. 9:1375-1382(1990).			
CC	-!- FUNCTION: INVOLVED IN THE SECRETION OF PROTEASES A, B, C AND G.			
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.			
CC	-!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY			
CC	(ABC TRANSPORTERS). HLYB SUBFAMILY.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M60395; AAA63634.1; -
EMBL; X53253; CAA37342.1; -
PIR; S12525; S12525.
HSP; F13569; 1NED.
PROSITE: PS00211; ABC_TRANSPORTER; 1.
PFAM; PF00005; ABC_tran; 1.
Transmembrane; Inner membrane.
Transmembrane; 22 42
Transmembrane; 56 76
Transmembrane; 133 153
Transmembrane; 154 174
Transmembrane; 255 275
Transmembrane; 364 371
Transmembrane; 575 AA; 61617 MW; 1BCCE4E74B1B92D6 CRC64;
SEQUENCE

Query Match 82.8%; Score 48; DB 1; Length 575;
 Best Local Similarity 77.8%; Pred. No. 9.30e+01;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 244 ASKYSRNL 252
 ||||| I
 QY 48 ASKYSRNL 56

RESULT 2
 ID DCE_PETHY STANDARD; PRT; 500 AA.
 AC Q07346;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE GLUTAMATE DECARBOXYLASE (EC 4.1.1.15) (GAD).
 GN GAD.
 OS Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
 OC Petunia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PETAL;
 RX MEDLINE; 93374956.
 RA Baum G., Chen Y., Arazi T., Takatsui H., Fromm H.;
 RT "A plant glutamate decarboxylase containing a calmodulin binding
 domain. Cloning, sequence, and functional analysis.";
 RL J. Biol. Chem. 268:19610-19617(1993).
 CC -!- FUNCTION: CATALYZES THE PRODUCTION OF GABA. THE CALMODULIN-BINDING
 IS CALCIUM-DEPENDENT AND IT IS PROPOSED THAT THIS WAY, DIRECTLY
 OR INDIRECTLY, FORM A CALCIUM REGULATED CONTROL OF GABA
 BIOSYNTHESIS.
 CC -!- CATALYTIC ACTIVITY: L-GLUTAMATE = 4-AMINOBUTANOATE + CO(2).
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -!- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND
 TYRDC).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: LI6797; AAA33709.1;
 DR EMBL; LI6797; AAA33710.1;
 DR PROSITE; PS00392; DDC_GAD_HDC_YDC; FALSE_NEG.
 DR PFAM; PF00282; Pyridoxal_gdc; 1.
 KW Lyase; Decarboxylase; Pyridoxal_gdc; 1.
 FT BINDING 277 277 PYRIDOXAL PHOSPHATE; Calmodulin-binding;
 FT DOMAIN 469 500 CALMODULIN-BINDING.
 SQ SEQUENCE 500 AA; 56726 MW; 72A043CB885AE10D CRC64;

Query Match 77.6%; Score 45; DB 1; Length 500;
 Best Local Similarity 66.7%; Pred. No. 5.24e+00;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 19 ASRYVRTSL 27
 ||| ||| I
 QY 48 ASKYSRNL 56

RESULT 3
 ID DCE1_ARATH STANDARD; PRT; 502 AA.
 AC Q42521;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GLUTAMATE DECARBOXYLASE 1 (EC 4.1.1.15) (GAD 1).
 GN GAD1 OR GDH1 OR GAD.

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; euroids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLOMBIA;
 RX MEDLINE; 95334488.
 RA Arazi T., Baum G., Snedden W.A., Shelp B.J., Fromm H.;
 RT "Molecular and biochemical analysis of calmodulin interactions with
 the calmodulin-binding domain of plant glutamate decarboxylase.";
 RL Plant Physiol. 108:551-561(1995).
 CC -!- FUNCTION: CATALYZES THE PRODUCTION OF GABA. THE CALMODULIN-BINDING
 IS CALCIUM-DEPENDENT AND IT IS PROPOSED THAT THIS WAY, DIRECTLY
 OR INDIRECTLY, FORM A CALCIUM REGULATED CONTROL OF GABA
 BIOSYNTHESIS.
 CC -!- CATALYTIC ACTIVITY: L-GLUTAMATE = 4-AMINOBUTANOATE + CO(2).
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -!- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND
 TYRDC).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: U10034; AAA93132.1;
 DR PROSITE; PS00392; DDC_GAD_HDC_YDC; FALSE_NEG.
 DR PFAM; PF00282; Pyridoxal_gdc; 1.
 KW Lyase; Decarboxylase; Pyridoxal_gdc; 1.
 FT BINDING 277 277 PYRIDOXAL PHOSPHATE; Calmodulin-binding;
 FT DOMAIN 469 502 CALMODULIN-BINDING.
 SQ SEQUENCE 502 AA; 57110 MW; FFC54636B5A2BFF8 CRC64;

Query Match 77.6%; Score 45; DB 1; Length 502;
 Best Local Similarity 66.7%; Pred. No. 5.24e+00;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 19 ASRYVRTSL 27
 ||| ||| I
 QY 48 ASKYSRNL 56

RESULT 4
 ID CATA_DEIRA STANDARD; PRT; 536 AA.
 AC Q59337;
 DT 15-FEB-2000 (Rel. 39, Created)
 DT 15-FEB-2000 (Rel. 39, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE CATALASE (EC 1.11.1.6).
 GN KATA.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-R1;
 RA Narumi I., Watanabe H., Hossain A., Tanaka A., Kitayama S.;
 RT "Molecular cloning and nucleotide sequence of radiation-inducible
 catalase gene from radioresistant bacterium, Deinococcus
 radiodurans.";
 RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN. SERVES
 TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.
 CC -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
 CC -!- COFACTOR: HEME GROUP.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
 CC -!- SIMILARITY: BELONGS TO THE CATALASE FAMILY.

US-09-376-430-2-10.rsp

Thu May 11 06:49:47 2000

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; D63898; BAA09937.1; --
CC HSP; P00432; 4BLC.
CC PRM; PF00199; catalase; 1.
CC PRINTS: PR00067; CATALASE.
CC PROSITE: PS00437; CATALASE_1; 1.
CC Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
CC ACT_SITE 81
CC ACT_SITE 159
CC BINDING 369
CC SEQUENCE 536 AA; 60622 MW; EE71985389DC682C CRC64;
CC

Query Match 77.6%; Score 45; DB 1; Length 536;
Best Local Similarity 62.5%; Pred. No. 5.24e+00; Indels 0; Gaps 0;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 103 SKYTRAKL 110
QY 49 SKYSRTNL 56
||||:|

RESULT 5 STANDARD; PRT; 789 AA.
ID R1RL_HSV4
AC P0642; 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE LARGE CHAIN (EC 1.17.4.1)
DE (RIBONUCLEOSIDE REDUCTASE).
OS Equine herpesvirus type 4 (strain 1942) (EHV-4) (Equine herpesvirus
OS type 1 subtype 2).
OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94266155.
RA Rigio M.P., Onions D.E.;
RT Sequences of the ribonucleotide reductase-encoding genes of equine
RT herpesvirus 4;
RL Gene 143:217-222(1994).
CC -!- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
CC -!- CATALYTIC ACTIVITY: 2'-DEOXYRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
CC -!- THIOREDUXIN + H(2)O - RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
CC THIOREDUXIN.
CC -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC LARGE CHAIN FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; X75354; CAA53100.1; --
CC PRINTS: PR01183; RIBORDTASEMI.
CC PROSITE: PS00089; RIBORED_LARGE; 1.
CC PFAM: PF00317; ribonucleo_red; 1.
CC Oxidoreductase; DNA replication; Early protein.
KW Oxidoreductase; DNA replication; 7CD9A24F3147988B CRC64;
SQ SEQUENCE 789 AA; 88046 MW; 88046 MW; 7CD9A24F3147988B CRC64;

Query Match 75.9%; Score 44; DB 1; Length 789;
Best Local Similarity 75.0%; Pred. No. 9.13e+00;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 575 SKYRGEL 582
QY 49 SKYSRTNL 56
||||:|

RESULT 6 STANDARD; PRT; 1551 AA.
ID YQ12_CAEEL
AC Q09449;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 175.7 KD PROTEIN C05C10.2 IN CHROMOSOME II.
DE C05C10.2.
OS Caenorhabditis elegans.
OS Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-BRISTOL N2;
RC Matthews P.;
RA Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RL -!- SIMILARITY: BELONGS TO THE DNA2/NAM7 HELICASE FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; Z48178; CAA89201.1; --
CC WORMPEP; C05C10.2; CE01466.
DR Hypothetical protein; ATP-binding; Helicase.
KW NP_BIND 1021 1028 ATP (BY SIMILARITY).
FT SEQUENCE 1551 AA; 175699 MW; 9C467FF9550104D6 CRC64;
SQ

Query Match 75.9%; Score 44; DB 1; Length 1551;
Best Local Similarity 55.8%; Pred. No. 9.13e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 16 SAKFSRNL 24
QY 48 ASKYSRTNL 56
||||:|

RESULT 7 STANDARD; PRT; 185 AA.
ID PSFAF_ODOSI
AC P49483;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE PHOTOSYSTEM I REACTION CENTRE SUBUNIT III (PSI-F).
DE PSFAF.
OS Odontella sinensis.
OG Chloroplast.
OC Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC Bidulphiophycidae; Eupodiscales; Odontella.
[1]
RN SEQUENCE FROM N.A.
RP Kowalik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U.;
RT "The chloroplast genome of a chlorophyll a-c-containing alga,
RT Odontella sinensis".
RL Plant Mol. Biol. Rep. 13:336-342(1995).
CC -!- FUNCTION: PROBABLY PARTICIPATES IN EFFICIENCY OF ELECTRON TRANSFER
CC FROM PLASTOCYANIN TO P700 (OR CYTOCHROME C553 IN ALGAE AND
CC CYANOBACTERIA). THIS PLASTOCYANIN-DOCKING PROTEIN CONTRIBUTES
CC TO THE SPECIFIC ASSOCIATION OF PLASTOCYANIN TO PS I.
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH LUMINAL SIDE OF THE
CC THYLAKOID MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE PSFAF FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; X75354; CAA53100.1; --
CC PRINTS: PR01183; RIBORDTASEMI.
CC PROSITE: PS00089; RIBORED_LARGE; 1.
CC PFAM: PF00317; ribonucleo_red; 1.
CC Oxidoreductase; DNA replication; Early protein.
KW Oxidoreductase; DNA replication; 7CD9A24F3147988B CRC64;
SQ SEQUENCE 789 AA; 88046 MW; 88046 MW; 7CD9A24F3147988B CRC64;

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: 267753; CAA91702.1; -
 DR MENDEL; 13331; ODO5; PsaF; 1;
 KW Photosynthesis; Photosystem I; Chloroplast; Thylakoid membrane.
 SQ SEQUENCE 185 AA; 20747 MW; 16C70038FAD4F048 CRC64;

Query Match 74.1%; Score 43; DB 1; Length 185;
 Best Local Similarity 71.4%; Pred. No. 1.58e+01;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 79 KYRSSEL 85
 |||||
 QY 50 KYSTRNL 56

RESULT 8
 ID VP2_HPV58 STANDARD; PRT; 358 AA.
 AC P26546;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE REGULATORY PROTEIN E2.
 GN E2.
 OS Human papillomavirus type 58.
 OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 9204102.
 RA Kiril Y., Iwamoto S., Matsukura T.;
 RT "Human papillomavirus type 58 DNA sequence.";
 RL Virology 185:424-427(1991).
 CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.

CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACNNNNNGGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -!- SUBUNIT: BINDS DNA AS A DIMER.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: D90400; BAA31848.1; -
 DR PIR; B36779; W2WL58.
 DR HSP; P17383; 1DHM.
 DR PFAM; PF00508; E2_N; 1.
 DR PFAM; PF00511; E2_C; 1.
 KW Early protein; Transcription regulation; Activator; DNA-binding;
 KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 SQ SEQUENCE 358 AA; 40781 MW; A3CA0BE001E2BD1E CRC64;

Query Match 74.1%; Score 43; DB 1; Length 358;
 Best Local Similarity 66.7%; Pred. No. 1.58e+01;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 175 AKYSKTQL 183
 |||||

QY 48 ASKYSRTNL 56

RESULT 9

ID KAA8_SCHPO STANDARD; PRT; 513 AA.
 AC O09792;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PROBABLE SERINE/THREONINE-PROTEIN KINASE C22G7.08 (EC 2.7.1.-).
 GN SPAC22G7.08.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
 OC Schizosaccharomycetaceae; Schizosaccharomyces.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA Badcock K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (OCT-1995) to the EMBL/Genbank/DBJ databases.
 CC -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
 CC PROTEIN KINASES

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: Z54328; CAA91132.1; -
 CC HSP; Q63450; IAO6.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PFAM; PF00069; pkinase; 1.
 KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;
 KW ATP-binding.

FT DOMAIN 241 505
 FT NP_BIND 247 255
 FT BINDING 270 270
 FT ACT_SITE 364 364
 SQ SEQUENCE 513 AA; 58546 MW; 61A6C73C5AFBDDDEA CRC64;

Query Match 74.1%; Score 43; DB 1; Length 513;
 Best Local Similarity 71.4%; Pred. No. 1.58e+01;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 44 KYRSSEL 50
 |||||

QY 50 KYSTRNL 56

RESULT 10

ID TFB1_YEAST STANDARD; PRT; 642 AA.
 AC P32776;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE RNA POLYMERASE II TRANSCRIPTION FACTOR B 73 KD SUBUNIT.
 GN TFB1 OR YDR311W OR D9740.3
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.

RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 314-322 AND 630-639.
 RX MEDLINE; 92410301.
 RA Gileadi O., Feaver W.J., Kornberg R.D.;
 RT "Cloning of a subunit of yeast RNA polymerase II transcription factor
 RT b and CTD kinase.";
 RL Science 257:1389-1392(1992).
 RN [2]
 RP SEQUENCE FROM N.A.

US-09-376-430-2-10.rsp

Thu May 11 06:49:47 2000

RC STRAIN-S288C / AB972;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favella A., Fulton L., Gattung S., Greco F., Kirsten J.,
 RA Kucaba T., Hallsworth K., Langston Y., Latreille P., Le T.,
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
 RA Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
 RA Rifken L., Riles L., Raich A., Trevaskis E., Vignati D.,
 RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.,
 RA Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RL FUNCTION: FACTOR B HAS CTD KINASE ACTIVITY AND DNA-DEPENDENT
 CC ATPASE ACTIVITY, AND IS ESSENTIAL FOR POLYMERASE II TRANSCRIPTION
 CC IN VITRO.
 CC SUBUNIT: COMPOSED OF SUBUNITS OF 85 KD, 73 KD AND 50 KD.
 CC SUBCELLULAR LOCATION: NUCLEAR.
 CC SIMILARITY: SOME, WITH MAMMALIAN BTF2.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M95750; AAA35143.1; -;
 CC EMBL; U28374; AAB64747.1; -;
 CC PIR; S31285; S31285.
 CC TRANSPAC; T02190; -;
 CC SCD; L0002284; TFB1.
 CC Transcription regulation; Nuclear protein.
 CC KW SEQUENCE 642 AA; 72894 MW; DAE369FE90A224F8 CRC64;
 CC
 CC Query Match 74.1%; Score 43; DB 1; Length 642;
 CC Best Local Similarity 100.0%; Pred. No. 1.58e+01;
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 Db 393 YSRNTL 398
 QY 51 YSRNTL 56
 PRT; 857 AA.
 RESULT 11
 ID NFM CHICK STANDARD; PRT; 857 AA.
 AC P16053; 1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-APR-1990 (Rel. 34, Last annotation update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE NEUROFILAMENT TRIPLET M PROTEIN (160 KD NEUROFILAMENT PROTEIN) (NF-M).
 GN NEFM.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
 CC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 CC
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 90174973.
 RA Zopf D., Dineva B., Betz H., Gundelfinger E.D., Betz H.;
 RT "Isolation of the chicken middle-molecular weight neurofilament
 RT (NF-M) gene and characterization of its promoter."
 RL Nucleic Acids Res. 18:521-529(1990).
 CC
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 88112814.
 RA Zopf D., Hermans-Borgmeyer I., Gundelfinger E.D., Betz H.;
 RT "Identification of gene products expressed in the developing chick
 RT visual system: characterization of a middle-molecular-weight
 RT neurofilament cDNA."
 RL Genes Dev. 1:699-708(1987).
 CC
 CC FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC
 CC PHOSPHORYLATED ON A NUMBER OF REPEATS OF THE TRIPLET K-S-P, NFM IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC
 CC SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X17102; CAA34958.1; -;
 CC EMBL; X05558; CAA29073.1; -;
 CC PIR; A27040; A27040.
 CC PIR; S08061; S08061.
 CC PIR; S15762; S15762.
 CC PROSITE; PS00226; IF, 1.
 CC PFAM; PF00038; filament; 1.
 CC Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone;
 CC KW Phosphorylation; Glycoprotein.
 CC INIT_MET 0
 CC FT DOMAIN 1 98
 CC FT ROD.
 CC FT TAIL.
 CC FT COIL 1A.
 CC FT LINKER 1.
 CC FT COIL 1B.
 CC FT LINKER 12.
 CC FT COIL 2A.
 CC FT LINKER 2.
 CC FT COIL 2B.
 CC FT DOMAIN 286 406
 CC FT CARBOHYD 46 46
 CC FT CARBOHYD 426 426
 CC FT CONFLICT 546 546
 CC FT SEQUENCE 857 AA; 95704 MW; 4E2E0FC6AC54778B CRC64;
 CC
 CC Query Match 74.1%; Score 43; DB 1; Length 857;
 CC Best Local Similarity 66.7%; Pred. No. 1.58e+01;
 CC Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 CC
 Db 48 SSSYKRTNL 56
 QY 48 ASKYRTNL 56
 PRT; 903 AA.
 RESULT 12
 ID VGLB_HSVIF STANDARD; PRT; 903 AA.
 AC P06436;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE GLYCOPROTEIN B PRECURSOR.
 GN GB OR UL27.
 OS Herpes simplex virus (type 1 / strain F).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 CC Alphaherpesvirinae; Simplexvirus.
 CC
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 85083254.
 RA Pellett P.E., Kousoulas K.G., Pereira L., Roizman B.;
 RT "Anatomy of the herpes simplex virus 1 strain F glycoprotein B gene:
 RT primary sequence and predicted protein structure of the wild type and
 RT of monoclonal-antibody-resistant mutants."
 RL J. Virol. 53:243-253(1985).
 CC
 RN SEQUENCE OF 1-176 FROM N.A.
 RP MEDLINE; 88306232.
 RA Hamerschmidt W., Conraths F., Mankertz J., Buhk H.-J., Pauli G.,
 RA Ludwig H.;

RT *Common epitopes of glycoprotein B map within the major DNA-binding
 RT proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex
 RL virus type 1 (HSV-1)."
 CC Virology 155:406-418(1988).
 CC -1- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.
 CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,
 CC GB, GC, GG, GD, GI, AND GE.
 CC -1- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE
 CC REQUIRED FOR VIRAL GROWTH.
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M14164; AAA45776.1; -
 DR EMBL; M21633; AAA45788.1; -
 DR PIR; A03750; VGBE1.
 DR PFAM; PF00606; Glycoprotein_B; 1.
 KW Glycoprotein; Transmembrane; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 903
 FT DOMAIN 31 729 GLYCOPROTEIN B.
 FT TRANSMEM 730 745 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 751 770 POTENTIAL.
 FT TRANSMEM 774 794 POTENTIAL.
 FT DOMAIN 795 903 POTENTIAL.
 FT CARBOHYD 86 86 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 140 140 POTENTIAL.
 FT CARBOHYD 397 397 POTENTIAL.
 FT CARBOHYD 429 429 POTENTIAL.
 FT CARBOHYD 488 488 POTENTIAL.
 FT CARBOHYD 673 673 POTENTIAL.
 SQ SEQUENCE 903 AA; 100104 MW; 73BDCA7813DB35E8 CRC64;

Query Match 74.1%; Score 43; DB 1; Length 903;
 Best Local Similarity 55.6%; Pred. No. 1.58e+01;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 209 TAKYVRNRL 217
 QY :||| |||
 48 ASKYSRTNL 56

RESULT 13
 ID VGLB_HSV1K STANDARD; PRT; 904 AA.
 AC P06437;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE GLYCOPROTEIN B PRECURSOR.
 GN GB OR UL27.
 OS Herpes simplex virus (type 1 / strain KOS).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Simplexvirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 84174058.
 RA Bzik D.J., Fox B.A., Deluca N.A., Person S.;
 RT "Nucleotide sequence specifying the glycoprotein gene, gB, of herpes
 RL simplex virus type 1,"
 RL Virology 133:301-314(1984).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RX MEDLINE; 87071654.
 RA Bzik D.J., Debroy C., Fox B.A., Pederson N.E., Person S.;
 RT "The nucleotide sequence of the gB glycoprotein gene of HSV-2 and
 RL comparison with the corresponding gene of HSV-1,"
 RL Virology 155:322-333(1986).

REVISIONS.
 RA Pederson N.E.;
 RL Submitted (APR-1987) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.
 CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,
 CC GB, GC, GG, GD, GI, AND GE.
 CC -1- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE
 CC REQUIRED FOR VIRAL GROWTH.
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; K01760; AAA45774.1; -
 DR PIR; A03751; VGBE1.
 DR PFAM; PF00606; Glycoprotein_B; 1.
 KW Glycoprotein; Transmembrane; Signal.
 FT SIGNAL 1 30
 FT CHAIN 31 904 GLYCOPROTEIN B.
 FT DOMAIN 31 730 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 731 746 POTENTIAL.
 FT TRANSMEM 752 771 POTENTIAL.
 FT TRANSMEM 775 795 POTENTIAL.
 FT DOMAIN 796 904 POTENTIAL.
 FT CARBOHYD 87 87 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 141 141 POTENTIAL.
 FT CARBOHYD 398 398 POTENTIAL.
 FT CARBOHYD 430 430 POTENTIAL.
 FT CARBOHYD 489 489 POTENTIAL.
 FT CARBOHYD 674 674 POTENTIAL.
 SQ SEQUENCE 904 AA; 100368 MW; B97B7F8DE5FBA299 CRC64;

Query Match 74.1%; Score 43; DB 1; Length 904;
 Best Local Similarity 55.6%; Pred. No. 1.58e+01;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 210 TAKYVRNRL 218
 QY :||| |||
 48 ASKYSRTNL 56

RESULT 14
 ID VGLB_HSV1P STANDARD; PRT; 904 AA.
 AC P08665;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE GLYCOPROTEIN B PRECURSOR.
 GN GB OR UL27.
 OS Herpes simplex virus (type 1 / strain Patton).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Simplexvirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87112925.
 RA Stuve L.L., Brown-Shimer S., Pacht C., Najarian R., Dina D.,
 RA Burke R.L.;
 RT "Structure and expression of the herpes simplex virus type 2
 RL glycoprotein gB gene,"
 RL J. Virol. 61:326-335(1987).
 CC -1- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.
 CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,
 CC GB, GC, GG, GD, GI, AND GE.
 CC -1- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE
 CC REQUIRED FOR VIRAL GROWTH.
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN B.
 CC -----

US-09-376-430-2-10.RSP

Thu May 11 06:49:47 2000

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL: K03541; AAA45778.1; -
 CC PFAM: PF00606; Glycoprotein_B; 1.
 KW Glycoprotein; Transmembrane; Signal.
 FT SIGNAL 1 30
 FT CHAIN 31 904
 FT DOMAIN 31 730
 FT TRANSMEM 731 746
 FT TRANSMEM 752 772
 FT TRANSMEM 775 795
 FT TRANSMEM 796 904
 FT DOMAIN 87 904
 FT CARBOHYD 87 141
 FT CARBOHYD 141 141
 FT CARBOHYD 141 398
 FT CARBOHYD 398 430
 FT CARBOHYD 430 489
 FT CARBOHYD 489 674
 FT CARBOHYD 674 674
 FT SEQUENCE 904 AA; 100115 MW; 7825ELDC830A626F CRC64;
 SQ

Query Match 74.1%; Score 43; DB 1; Length 904;
 Best Local Similarity 55.6%; Pred. No. 1.58e+01;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 210 TAKYVRNRL 218
 : : : : :
 Qy 48 ASKYSRTNL 56

RESULT 15 STANDARD; PRT; 904 AA.
 ID VGLB_HSV11
 AC P10211; 1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE GLYCOPROTEIN B PRECURSOR.
 GN GB OR UL27.
 OS Herpes simplex virus (type 1 / strain 17).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Simplexvirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 88274327.
 RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
 RA McNab D., Perry L.J., Scott J.E., Taylor P.;
 RA "The complete DNA sequence of the long unique region in the genome of
 RT herpes simplex virus type 1."
 RL J. Gen. Virol. 69:1531-1574 (1988).
 CC -|- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.
 CC -|- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,
 CC GB, GC, GG, GI, AND GE.
 CC -|- REQUIRED FOR VIRAL GROWTH
 CC -|- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN B.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D10879; BAA01673.1; -
 CC EMBL: X14112; CAA32320.1; -
 CC FIR: I30084; VGBEW.
 CC PFAM: PF00606; Glycoprotein_B; 1.

KW Glycoprotein; Transmembrane; Signal.
 FT SIGNAL 1 30
 FT CHAIN 31 904
 FT DOMAIN 31 730
 FT TRANSMEM 731 746
 FT TRANSMEM 752 771
 FT TRANSMEM 775 795
 FT TRANSMEM 796 904
 FT DOMAIN 87 904
 FT CARBOHYD 87 141
 FT CARBOHYD 141 141
 FT CARBOHYD 141 398
 FT CARBOHYD 398 430
 FT CARBOHYD 430 489
 FT CARBOHYD 489 674
 FT CARBOHYD 674 674
 FT SEQUENCE 904 AA; 100292 MW; 2C14E8B1284C1F3A CRC64;
 SQ

Query Match 74.1%; Score 43; DB 1; Length 904;
 Best Local Similarity 55.6%; Pred. No. 1.58e+01;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 210 TAKYVRNRL 218
 : : : : :
 Qy 48 ASKYSRTNL 56

Search completed: Wed May 10 12:21:01 2000
 Job time : 99 secs.

Thu May 11 06:50:20 2000

 M P S R L H
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed May 10 14:15:32 2000; MasPar time 89.68 seconds
 2.377 Million cell updates/sec

Tabular output not generated.

Title: >US-09-376-430-2
 Description: (347-353) from US09376430A.ppt (24 of 25)
 Perfect Score: 50
 Sequence: 1 PLQGGDV 7

Scoring table: PAM 150
 Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot38
 1:swissprot

Statistics: Mean 20.757; Variance 18.906; scale 1.098

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	45	90.0	164	1	YOL6_CABEL	7.05e+01
2	44	88.0	209	1	PROLACTIN PRECURSOR (P	1.32e+00
3	44	88.0	750	1	HYPE_ECOLI	1.32e+00
4	43	86.0	394	1	HYF1_ALCEU	2.43e+00
5	43	86.0	424	1	VP3_BPHK7	2.43e+00
6	43	86.0	447	1	AFOA_SYNY3	2.43e+00
7	41	82.0	115	1	NUFM_RAT	8.04e+00
8	41	82.0	284	1	TRBG_AGR16	8.04e+00
9	41	82.0	291	1	AMPM_CHLTR	8.04e+00
10	41	82.0	334	1	YCL2_ECOLI	8.04e+00
11	41	82.0	379	1	Y528_SYNY3	8.04e+00
12	41	82.0	578	1	PIEF_RHOCA	8.04e+00
13	41	82.0	2411	1	MYSA_DROME	8.04e+00
14	41	82.0	2768	1	THYG_MOUSE	8.04e+00
15	40	80.0	225	1	ACE1_YEAST	1.44e+01
16	40	80.0	394	1	AMPL_HUMAN	1.44e+01
17	40	80.0	408	1	NK1R_RANCA	1.44e+01
18	40	80.0	408	1	ARCA_MYCHO	1.44e+01
19	40	80.0	525	1	PEN3_ADEG1	1.44e+01
20	40	80.0	524	1	FER3_CANAL	1.44e+01
21	40	80.0	692	1	YK06_YEAST	1.44e+01
22	40	80.0	1451	1	SPT6_YEAST	1.44e+01
23	40	80.0	2051	1	FAS1_YEAST	1.44e+01

24	39	78.0	156	1	KDTB_HAELN	LIPOLYSACCHARIDE COR	2.54e+01
25	39	78.0	159	1	FAB2_CABEL	FATTY ACID-BINDING PRO	2.54e+01
26	39	78.0	170	1	MLR1_CABEL	MYOSIN REGULATORY LIGH	2.54e+01
27	39	78.0	170	1	MLR2_CABEL	MYOSIN REGULATORY LIGH	2.54e+01
28	39	78.0	270	1	TRBG_RHLSN	PROBABLE CONJUGAL TRAN	2.54e+01
29	39	78.0	326	1	YA55_METJA	HYPOTHETICAL PROTEIN M	2.54e+01
30	39	78.0	367	1	RTCL1_YEAST	RNA 3'-TERMINAL PHOSPH	2.54e+01
31	39	78.0	454	1	CBPH_LOPAM	CARBOXYPEPTIDASE H PRE	2.54e+01
32	39	78.0	494	1	EXOT_RHIME	SUCCINOGLYCAN BIOSYNTH	2.54e+01
33	39	78.0	515	1	PEN3_ADEG1	PENTON PROTEIN (VIRION	2.54e+01
34	39	78.0	533	1	CAG2_HUMAN	BETA-1,4 N-ACETYLGLAC	2.54e+01
35	39	78.0	533	1	CAG2_RAT	BETA-1,4 N-ACETYLGLAC	2.54e+01
36	39	78.0	533	1	CAG2_MOUSE	BETA-1,4 N-ACETYLGLAC	2.54e+01
37	39	78.0	588	1	CGMA_RHIME	CYCLOC BETA-1,2-GLUCAN	2.54e+01
38	39	78.0	686	1	VGLH_PVNV3	GLYCOPROTEIN H PRECURS	2.54e+01
39	39	78.0	686	1	VGLH_PVRI	GLYCOPROTEIN H PRECURS	2.54e+01
40	39	78.0	686	1	VGLH_PVKA	GLYCOPROTEIN H PRECURS	2.54e+01
41	39	78.0	789	1	CLC7_HUMAN	CHLORIDE CHANNEL PROTE	2.54e+01
42	39	78.0	1102	1	YE20_METJA	HYPOTHETICAL PROTEIN M	2.54e+01
43	39	78.0	1379	1	MET_MOUSE	HEPATOCYTE GROWTH FACT	2.54e+01
44	38	76.0	274	1	YQ01_AERPE	HYPOTHETICAL PROTEIN A	4.42e+01
45	38	76.0	919	1	DNL1_HUMAN	DNA LIGASE I (EC 6.5.1	4.42e+01

ALIGNMENTS

RESULT	ID	YOL6_CABEL	STANDARD;	PRT;	164 AA.
AC	Q02333				
DT	01-FEB-1994	(Rel. 28, Created)			
DT	01-FEB-1994	(Rel. 28, Last sequence update)			
DT	01-JUN-1994	(Rel. 29, Last annotation update)			
DE	HYPOTHETICAL 18.4 KD PROTEIN ZK370.6 IN CHROMOSOME III.				
GN	ZK370.6				
OS	Caenorhabditis elegans.				
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;				
OC	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-BRISTOL N2;				
RX	MEDLINE: 94150718.				
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Kershaw J., Kirsten J., Laister N., Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhen R., Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden K., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Watson R., Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;				
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."				
RL	Nature 368:32-38(1994).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL: M98552; AAA28208.1;				
DR	WORMPEP; ZK370.6; CE00398.				
KW	Hypothetical protein.				
SK	SEQUENCE 164 AA; 18380 MW; B201A857F599CAB8 CRC64;				

Query Match 90.0%; Score 45; DB 1; Length 164;
 Best Local Similarity 85.7%; Pred. No. 7.05e+01;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Db 71 PLOGGV 77
QY 347 PLOGGDV 353

RESULT 2
ID PRL-ANGAN STANDARD; PRT; 209 AA.
AC P33096;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PROLACTIN PRECURSOR (PRL)
OS Anguilla anguilla (European freshwater eel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Anguilliformes; Anguillidae; Anguilla.
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=PIUITARY;
RX MEDLINE; 95011031.
RA Querat B., Cardinaud B., Hardy A., Vidal B., D'Angelo G.;
RT "Sequence and regulation of European eel prolactin mRNA.";
RL Mol. Cell. Endocrinol. 102:151-160(1994).
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X69149; CAA48902.1; -
DR PIR; S30541; S30541.
DR HSP; Q28632; 1AN3.
DR PRINTS; PR00836; SOMATOTROPIN.
DR PROSITE; PS00266; SOMATOTROPIN_1; 1.
DR PROSITE; PS00338; SOMATOTROPIN_2; 1.
DR PFAM; PF00103; hormone; 1.
DR Hormone; Pituitary; Signal.
FT SIGNAL 1 24 BY SIMILARITY.
FT CHAIN 25 209 PROLACTIN.
FT DISULFID 70 184 BY SIMILARITY.
FT DISULFID 201 209 BY SIMILARITY.
SQ SEQUENCE 209 AA; FC7B6FB67C79BB0 CRC64;

Query Match 88.0%; Score 44; DB 1; Length 209;
Best Local Similarity 85.7%; Pred. No. 1.32e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 161 PLOGGD 167
QY 347 PLOGGDV 353

RESULT 3
ID HYPF-ECOLI STANDARD; PRT; 750 AA.
AC P30131; Q46878;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYDROGENASE MATURATION PROTEIN HYPF.
GN HYPF OR HYDA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Tomiyama M., Shiotani M., Nishio M., Ikebukuro K., Sode K.,
RA Tamaya E., Karube I.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.

[2]
SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE; 97426617.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Klinkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
FUNCTION.
RX MEDLINE; 96241867.
RA Maier T., Binder U., Bock A.;
RT "Analysis of the hda locus of Escherichia coli: two genes (hdaN and
RT hdaP) involved in formate and hydrogen metabolism.";
RL Arch. Microbiol. 165:333-341(1996).
CC -!- FUNCTION: INVOLVED IN THE HYDROGENASES 1, 2 AND 3 MATURATION
CC PROCESS.
CC -!- SIMILARITY: BELONGS TO THE HYPF FAMILY.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC ACYLPHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D14422; BAA03315.1; -
DR EMBL; U29579; AAA69222.1; -
DR EMBL; AE000355; AAC75754.1; -
DR HSP; P41500; 2ACY.
DR ECGENE; EG11551; HYPF.
DR PROSITE; PS00150; ACYLPHOSPHATASE_1; FALSE_NEG.
DR PFAM; PF01300; Sua5_yc10_yc10; 1.
DR Zinc-finger.
FT DOMAIN 1 91 ACYLPHOSPHATASE-LIKE.
FT ZN_FING 109 134 C4-TYPE (POTENTIAL).
FT ZN_FING 159 184 C4-TYPE (POTENTIAL).
FT CONFLICT 31 32 QQ -> RE (IN REF. 1).
FT CONFLICT 367 386 SGEMLRSSRGVYDADALPP -> RNAAPPFAGVCAGCAGF
FT AS (IN REF. 1).
FT CONFLICT 431 432 WR -> CA (IN REF. 1).
FT CONFLICT 435 435 L -> S (IN REF. 1).
SQ SEQUENCE 750 AA; 82066 MW; 8B245F80349A91E CRC64;

Query Match 88.0%; Score 44; DB 1; Length 750;
Best Local Similarity 85.7%; Pred. No. 1.32e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 490 PLOGGDV 496
QY 347 PLOGGDV 353

RESULT 4
ID HYPF-ALCEU STANDARD; PRT; 394 AA.
AC P45805;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYDROGENASE MATURATION PROTEIN HYPF1.
GN HYPF1 OR HYPF.
OS Alcaligenes eutrophus.
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Ralstonia.
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=H16 / ATCC 17699;

```

```
RX MEDLINE; 93356597.
RA Darnedde J., Eitinger M., Friedrich B.;
RT "Analysis of a pleiotropic gene region involved in formation of
RL catalytically active hydrogenases in Alcaligenes eutrophus H16.";
RT Arch. Microbiol. 159:545-553(1993).
RN [2]
RP REVISIONS.
RC STRAIN-H16 / ATCC 17699;
RA Eitinger T.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE HYDROGENASE MATURATION PROCESS.
CC -!- SIMILARITY: BELONGS TO THE HYPF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X70183; CAA49731.1;
DR HSSP; P02647; 10DR.
KW Plasmid.
SQ SEQUENCE 394 AA; 40677 MW; 268E9CF0C9B3E64B CRC64;
Query Match 86.0%; Score 43; DB 1; Length 394;
Best Local Similarity 85.7%; Pred. No. 2.43e+00;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 157 PLPGGDV 163
QY 347 PLQGGDV 353
RESULT 5
ID VP3_BPHK7 STANDARD; PRT; 424 AA.
AC P49859;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PORTAL PROTEIN (GP3).
GN 3.
OS Bacteriophage HK97.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC Lambda phage group.
CC [1]
RN SEQUENCE FROM N.A.
RX Duda R.L., Martincic K., Hendrix R.W.;
RT "Genetic basis of bacteriophage HK97 prohead assembly.";
RL J. Mol. Biol. 247:636-647(1995).
CC -!- FUNCTION: HEAD-TAIL CONNECTOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U18319; AAA80202.1;
DR SEQUENCE 424 AA; 47266 MW; DD75FDCFD11947E1 CRC64;
Query Match 86.0%; Score 43; DB 1; Length 424;
Best Local Similarity 85.7%; Pred. No. 2.43e+00;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 394 PLPGGDV 400
QY 347 PLQGGDV 353
```

```
RESULT 6
ID AROA_SVNY3 STANDARD; PRT; 447 AA.
AC Q59975; Q59974;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19)
DE (3-ENOLPYRUVYLSHIKIMATE-5-PHOSPHATE SYNTHASE) (EPSP SYNTHASE).
GN AROA OR SLR0444.
OS Synchocystis sp. (strain PCC 6803)
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
CC [1]
CC SEQUENCE FROM N.A.
RX MEDLINE; 94299161.
RA dalla Chiesa M., Mayes S.R., Maskell D.S., Nixon P.J., Barber J.;
RT "An aroA homologue from Synchocystis sp. PCC 6803.";
RL Gene 144:145-146(1994).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 97061201.
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
RN [3]
RN SEQUENCE OF 103-194 FROM N.A.
RX MEDLINE; 93307506.
RA Mayes S.R., dalla Chiesa M., Zhang Z., Barber J.;
RT "The genes aroA and trnQ are located upstream of psbO in the
RT chromosome of Synchocystis 6803.";
RL FEBS Lett. 325:255-261(1993).
CC -!- CATALYTIC ACTIVITY: PHOSPHOENOLPYRUVATE + 3-PHOSPHOSHIKIMATE =
CC ORTHOPHOSPHATE + O(5)-(1-CARBOXYVINYL)-3-PHOSPHOSHIKIMATE.
CC -!- PATHWAY: SIXTH STEP IN THE BIOSYNTHESIS FROM CHORISMATE OF THE
CC AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).
CC -!- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X75325; CAA53074.1;
DR EMBL; D90914; BAA18477.1;
DR EMBL; X72784; CAA51291.1;
DR PROSITE; PS00104; EPSP-SYNTHASE_1; 1.
DR PROSITE; PS00885; EPSP-SYNTHASE_2; 1.
DR PFAM; PF00275; EPSP_synase; 1
KW Aromatic amino acid biosynthesis; Transferase.
SQ SEQUENCE 447 AA; 47046 MW; 2F3C8AD2685A7BCE CRC64;
Query Match 86.0%; Score 43; DB 1; Length 447;
Best Local Similarity 71.4%; Pred. No. 2.43e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 393 PLQGAEV 399
QY 347 PLQGGDV.353
RESULT 7
ID NUFM_RAT STANDARD; PRT; 115 AA.
AC Q63362;
DT 01-NOV-1997 (Rel. 35, Created)
```

01-NOV-1997 (Rel. 35, Last sequence update)
 15-JUL-1998 (Rel. 36, Last annotation update)
 NADH-UBIQUINONE OXIDOREDUCTASE 13 KD-B SUBUNIT (EC 1.6.5.3)
 (EC 1.6.99.3) (COMPLEX I-13KD-B) (B13).
 NDUFAS5.
 GN Rattus norvegicus (Rat).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 SEQUENCE FROM N.A.
 RP STRAIN-WISTAR; TISSUE-BRAIN;
 RC MEDLINE; 97029440.
 RX Kitahara T., Takeda N., Kubo T., Kiyama H.;
 RA "Molecular cloning of the rat NADH:ubiquinone oxidoreductase subunit
 RT and its up-regulation in the facial muscle after denervation;
 RT detected by means of differential display.";
 RL Neurol. Res. 18:329-336(1996).
 CC -!- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
 CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
 CC TO BE UBIQUINONE. THIS IS A COMPONENT OF THE IRON-SULFUR (IP)
 CC FRAGMENT OF THE ENZYME.
 CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
 CC -!- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE; MATRIX SIDE.
 CC -!- SIMILARITY: BELONGS TO THE COMPLEX I 29.9 KD SUBUNIT FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D86215; BAA13045.1; -
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Acetylation.
 FT INIT_MET 0 0 BY SIMILARITY.
 FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
 SQ SEQUENCE 115 AA; 13281 MW; B2689F269333A5DA CRC64;
 Query Match 82.0%; Score 41; DB 1; Length 115;
 Best Local Similarity 83.3%; Pred. No. 8.04e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 71 LOGGEV 76
 |||||
 QY 348 LOGGDV 353
 RESULT 8
 ID TRBG-AGRT6 STANDARD; PRT; 284 AA.
 AC P54915;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE CONJUGAL TRANSFER PROTEIN TRBG PRECURSOR.
 GN TRBG.
 OS Agrobacterium tumefaciens.
 OG Plasmid pTiA6NC.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Agrobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX Alt-Morbe J., Stryker J.L., Fuqua C., Li P.L., Farrand S.K.,
 RA Winans S.C.;
 RA "The conjugal transfer system of Agrobacterium tumefaciens
 RT octopine-type Ti plasmids is closely related to the transfer system
 RT of an IncP plasmid and distantly related to Ti plasmid vir genes.";
 RL J. Bacteriol. 178:4248-4257(1996).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U43675; AAB95102.1; -
 KW Conjugation; Plasmid; Signal.
 FT SIGNAL 1 34 POTENTIAL.
 FT CHAIN 35 284 CONJUGAL TRANSFER PROTEIN TRBG.
 SQ SEQUENCE 284 AA; 30934 MW; 6C753A64F31FB78C CRC64;
 Query Match 82.0%; Score 41; DB 1; Length 284;
 Best Local Similarity 83.3%; Pred. No. 8.04e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 90 LOGGEV 95
 |||||
 QY 348 LOGGDV 353
 RESULT 9
 ID AMPM-CHLTR STANDARD; PRT; 291 AA.
 AC O84859;
 DT 15-FEB-2000 (Rel. 39, Created)
 DT 15-FEB-2000 (Rel. 39, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE METHIONINE AMINOPEPTIDASE (EC 3.4.11.18) (MAP) (PEPTIDASE M).
 GN MAP OR CT851.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-D/UW-3/CX;
 RX MEDLINE; 99008089.
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RA "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 CC -!- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT
 CC PROTEINS (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: L-METHIONYLPEPTIDE + H(2)O = L-METHIONINE +
 CC PEPTIDE.
 CC -!- COFACTOR: COBALT (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24A; ALSO KNOWN AS THE
 CC MAP FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF001358; AAC68448.1; -
 DR PRINTS; PR00599; MAPEPTIDASE.
 DR PROSITE; PS00680; MAP_1; 1.
 DR PFAM; PF00557; Peptidase_M24; 1.
 KW Hydrolase; Amino-peptidase; Cobalt.
 FT METAL 135 135 COBALT (BY SIMILARITY).
 FT METAL 145 146 COBALT (BY SIMILARITY).
 FT METAL 209 209 COBALT (BY SIMILARITY).
 FT METAL 241 241 COBALT (BY SIMILARITY).
 FT METAL 274 274 COBALT (BY SIMILARITY).
 SQ SEQUENCE 291 AA; 32640 MW; DCC01BC0157COA89 CRC64;
 Query Match 82.0%; Score 41; DB 1; Length 291;
 Best Local Similarity 71.4%; Pred. No. 8.04e+00;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Thu May 11 06:50:20 2000

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; D64006; BAA10876.1; -
CC PROSITE; PS00142; ZINC-PROTEASE; UNKNOWN_1.
CC Hypothetical protein; Transmembrane.
CC TRANSMEM 20 40 POTENTIAL.
CC TRANSMEM 54 74 POTENTIAL.
CC TRANSMEM 115 135 POTENTIAL.
CC TRANSMEM 148 168 POTENTIAL.
CC TRANSMEM 212 232 POTENTIAL.
CC SEQUENCE 379 AA; 40465 MW; 1CC3251660078ECB CRC64;
CC -----

Query Match 82.0%; Score 41; DB 1; Length 379;
Best Local Similarity 71.4%; Pred. No. 8.04e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 169 PLDGGV 175
QY 347 PLQGGDV 353
||| |||
||:|:|

RESULT 12
ID PTFB_RHOCA STANDARD; PRT; 578 AA.
AC P23387;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE PTS SYSTEM, FRUCTOSE-SPECIFIC IIBC COMPONENT (EIIBC-FRU) (FRUCTOSE-
DE PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)
DE (EC 2.7.1.69) (EII-FRU).
GN FRUA
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=37B4;
RX MEDLINE; 91072273.
RA Wu L.-F., Saier M.H. Jr.;
RT "Nucleotide sequence of the fruA gene, encoding the fructose permease
RT of the Rhodobacter capsulatus phosphotransferase system, and analyses
RT of the deduced protein sequence."
RL J. Bacteriol. 172:7167-7178(1990).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS 2 PTS EIIIB DOMAINS.
CC -!- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; X53150; CAA37303.1; -
CC PIR; B37852; B37852.
CC Phosphotransferase system; Sugar transport; Transferase;
CC Transmembrane; Phosphorylation.
CC -----

CC EMBL; D64006; BAA10876.1; -
CC PROSITE; PS00142; ZINC-PROTEASE; UNKNOWN_1.
CC Hypothetical protein; Transmembrane.
CC TRANSMEM 20 40 POTENTIAL.
CC TRANSMEM 54 74 POTENTIAL.
CC TRANSMEM 115 135 POTENTIAL.
CC TRANSMEM 148 168 POTENTIAL.
CC TRANSMEM 212 232 POTENTIAL.
CC SEQUENCE 379 AA; 40465 MW; 1CC3251660078ECB CRC64;
CC -----

Query Match 82.0%; Score 41; DB 1; Length 379;
Best Local Similarity 71.4%; Pred. No. 8.04e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 169 PLDGGV 175
QY 347 PLQGGDV 353
||| |||
||:|:|

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; D64006; BAA10876.1; -
CC PROSITE; PS00142; ZINC-PROTEASE; UNKNOWN_1.
CC Hypothetical protein; Transmembrane.
CC TRANSMEM 20 40 POTENTIAL.
CC TRANSMEM 54 74 POTENTIAL.
CC TRANSMEM 115 135 POTENTIAL.
CC TRANSMEM 148 168 POTENTIAL.
CC TRANSMEM 212 232 POTENTIAL.
CC SEQUENCE 379 AA; 40465 MW; 1CC3251660078ECB CRC64;
CC -----

Query Match 82.0%; Score 41; DB 1; Length 379;
Best Local Similarity 71.4%; Pred. No. 8.04e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 169 PLDGGV 175
QY 347 PLQGGDV 353
||| |||
||:|:|

RESULT 11
ID Y528_SYNY3 STANDARD; PRT; 379 AA.
AC Q55518;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 40.5 KD PROTEIN SLL0528.
GN SLL0528.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96127529.
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome."
RL DNA Res. 2:153-166(1995).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: TO B.SUBTILIS SPOIVB AND TO M.JANNASCHII MJ0392.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; Z17241; CAA78940.1; -
CC PIR; S31733; S31733.
CC PIR; S33669; S33669.
CC PFAM; PF01370; Epimerase; 1.
KW Hypothetical protein; Lyase; NAD.
QY SEQUENCE 334 AA; 37601 MW; 3C6CD57B7F5FC02D CRC64;
CC -----

Query Match 82.0%; Score 41; DB 1; Length 334;
Best Local Similarity 85.7%; Pred. No. 8.04e+00;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 289 PLQGGDV 295
QY 347 PLQGGDV 353
||| |||
||:|:|

RESULT 11
ID Y528_SYNY3 STANDARD; PRT; 379 AA.
AC Q55518;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 40.5 KD PROTEIN SLL0528.
GN SLL0528.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96127529.
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome."
RL DNA Res. 2:153-166(1995).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: TO B.SUBTILIS SPOIVB AND TO M.JANNASCHII MJ0392.
CC -----

```

FT DOMAIN 1 93
FT DOMAIN 116 203
FT DOMAIN 125 578
FT MOD_RES 125 125
FT TRANSMEM 4 24
FT TRANSMEM 44 64
FT TRANSMEM 119 139
FT TRANSMEM 251 271
FT TRANSMEM 284 304
FT TRANSMEM 319 339
FT TRANSMEM 364 384
FT TRANSMEM 385 405
FT TRANSMEM 406 426
FT TRANSMEM 445 465
FT TRANSMEM 477 497
FT TRANSMEM 518 538
FT TRANSMEM 545 565
SQ SEQUENCE 578 AA; 58575 MW; 87793E5AAA08E4E CRC64;

Query Match 82.0%; Score 41; DB 1; Length 578;
Best Local Similarity 57.1%; Pred. No. 8.04e+00;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 48 PLOGAEI 54
    |||||
Qy 347 PLOGGDV 353

RESULT 13
ID MISA_DROME STANDARD; PRT; 2411 AA.
AC P05661;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MYOSIN HEAVY CHAIN, MUSCLE.
GN MHC.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89384556.
RA George E.L., Ober M.B., Emerson C.P. Jr.;
RT "Functional domains of the Drosophila melanogaster muscle myosin
heavy-chain gene are encoded by alternatively spliced exons.";
RL Mol. Cell. Biol. 9:2957-2974(1989).
RN [2]
RP SEQUENCE OF 1-312 FROM N.A.
RX MEDLINE; 87280141.
RA Wassenberg D.R. II, Kronert W.A., O'Donnell P.T., Bernstein S.I.;
RT "Analysis of the 5' end of the Drosophila muscle myosin heavy chain
gene. Alternatively spliced transcripts initiate at a single site and
intron locations are conserved compared to myosin genes of other
organisms.";
RL J. Biol. Chem. 262:10741-10747(1987).
RN [3]
RP SEQUENCE OF 486-881 FROM N.A.
RX STRAIN-CANTON-S; TISSUE-EMBRYONIC MUSCLE;
RX MEDLINE; 91330870.
RA Kronert W.A., Edwards K.A., Roche E.S., Wells L., Bernstein S.I.;
RT "Muscle-specific accumulation of Drosophila myosin heavy chains: a
splicing mutation in an alternative exon results in an isoform
substitution.";
RL EMBO J. 10:2479-2488(1991).
RN [4]
RP FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -1- ALTERNATIVE PRODUCTS: AS DROSOPHILA HAS A SINGLE MUSCLE MHC GENE,
MHC ISOFORMS MUST RELY ON ALTERNATIVE RNA SPLICING THAT LEADS TO
DIFFERENCES IN THE C-TERMINUS OF THE VARIOUS MHC PROTEINS.

```

-1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 -1- SIMILARITY: THE SEQUENCE SHOWS HIGH HOMOLOGY WITHIN THE PUTATIVE
 ATP BINDING DOMAIN, INDICATING A STRONG FUNCTIONAL CONSERVATION OF
 THIS DOMAIN AMONG THE VARIOUS MUSCLE AND NON-MUSCLE MHC PROTEINS.
 -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; M61229; AAA28686.1; ALT_SEQ.
 EMBL; M61229; AAA28687.1; ALT_SEQ.
 EMBL; J02788; AAA28706.1; ALT_SEQ.
 EMBL; J02788; AAA28707.1; ALT_SEQ.
 EMBL; X60196; CAA42752.1; ALT_SEQ.
 EMBL; X60196; CAA42753.1; ALT_SEQ.
 EMBL; X60196; CAA42754.1; ALT_SEQ.
 PIR; A28492; A28492.
 PIR; A32491; A32491.
 PIR; B32491; B32491.
 HSSP; P08799; LMND.
 DR FLYBASE; FBgn0002741; Mhc.
 DR PFAM; PF00063; myosin_head; 10.
 DR PFAM; PF00612; IQ; 1.
 DR PFAM; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Alternative splicing; Multigene family;
 KW Heptad repeat pattern.
 FT DOMAIN 1225 2376 COILED COIL (POTENTIAL).
 FT NP_BIND 227 234 ATP (BY SIMILARITY).
 FT VARSPLIC 2385 2385 P -> I (IN SHORT ISOFORM).
 FT VARSPLIC 2386 2411 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 43 44 EK -> RE (IN REF. 2).
 SQ SEQUENCE 2411 AA; 276434 MW; 8C4FEC1BB8D310A9 CRC64;
 Query Match 82.0%; Score 41; DB 1; Length 2411;
 Best Local Similarity 83.3%; Pred. No. 8.04e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 64 LQGEV 69
 Qy 348 LQGGDV 353
 |||||
 RESULT 14
 ID THYG_MOUSE STANDARD; PRT; 2768 AA.
 AC O08710;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE THYROGLOBULIN PRECURSOR.
 GN TG OR TGN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 98008184.
 RA Catuogre P., Vidalain P.O., Vali M., Aguilera-Galaviz L.A.,
 RA Rose N.R.;
 RT "Cloning and characterization of murine thyroglobulin cDNA";
 RL Clin. Immunol. Immunopathol. 85:221-226(1997).
 CC -1- FUNCTION: PRECURSOR OF THE IODINATED THYROID HORMONES THYROXINE
 (T4) AND TRIIODOTHYRONINE (T3).
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).

DR EMBL; M24390; AAA34386.1; -
 DR EMBL; Z48618; CAA88533.1; -
 DR EMBL; X85757; CAA59765.1; -
 DR EMBL; M28520; AAA34542.1; -
 DR EMBL; Z72688; CAA96877.1; -
 DR PIR; A31926; A31926; -
 DR PIR; A30133; A30133; -
 DR TRANSFAC; T00179; -
 DR SGD; L0000439; CUP2.
 DR PRINTS; PR00617; COPPERFIST.
 DR PROSITE; PS01119; COPPER_FIST_1; 1.
 DR PROSITE; PS50073; COPPER_FIST_2; 1.
 DR PFAM; PF00649; Copper-fist; 1.
 KW Transcription regulation; DNA-binding; Activator; Metal-binding;
 FT DNA_BIND 1 40 BINDS COPPER AND DNA.
 FT DOMAIN 1 108 BINDS COPPER AND DNA.
 FT DOMAIN 109 225 REQUIRED FOR TRANSCRIPTIONAL ACTIVATION.
 SQ SEQUENCE 225 AA; 24425 MW; E3199A3C8CA9CC1B CRC64;

Query Match 80.0%; Score 40; DB 1; Length 225;
 Best Local Similarity 57.1%; Pred. No. 1.44e+01;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 149 PLONGEI 155
 QY 347 PLOGGDV 353

Search completed: Wed May 10 14:17:11 2000
 Job time : 99 secs.

RA MYLER P.J., AUDLEMAN L., HIXSON G., KISER P., LEMLEY C., RICKEL E.,
 RA SISK E., SUNKIN S., SWARTZELL S., WESTLAKE T., MAGNESS C., BASTIEN P.,
 RA FU G., IVENS A., STUART K.;
 RT "Leishmania major Friedlin chromosome 1 has only two polycistronic
 RL units of protein coding genes";
 RN Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN-FRIEDLIN;
 RA MYLER P.J.;
 RT "The nucleotide sequence of Leishmania major Friedlin chromosome 1";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE001274; AAC24660.1; -;
 DR HSP; P08653; ILIC.
 SQ SEQUENCE 1607 AA; 168973 MW; F2354CA0 CRC32;

Query Match 88.0%; Score 44; DB 5; Length 1607;
 Best Local Similarity 85.7%; Pred. No. 2.15e+00;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 363 PLOGGSDV 369
 QY 347 PLOGGSDV 353
 |||||

RESULT 3
 ID P94996 PRELIMINARY; PRT; 2126 AA.
 AC 061795;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DE HYPOTHETICAL 221.0 KD PROTEIN.
 GN MYC06H11.26.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RA MURPHY L., HARRIS D.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RA BARRELL B.G., RAJANDREAM M.A.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE; 96161548.
 RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
 RA COLE S.T.;
 RT "An integrated map of the genome of the tubercle bacillus,
 RT Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium
 RT leprae";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
 DR EMBL; 285982; CAB06632.1; -;
 DR PROSITE; PS00012; PHOSPHATANTHINE; 1.
 DR PROSITE; PS00606; B-KETOACYL SYNTHASE; 1.
 DR PFAM; PF00698; Acyl_transf; 1.
 DR PFAM; PF00107; adh_zinc; 1.
 DR PFAM; PF00109; ketoacyl-synt; 1.
 DR PFAM; PF00550; pp-binding; 1.
 DR Hypothetical protein; Transferrase.
 SQ SEQUENCE 2126 AA; 221045 MW; ED3B9C16 CRC32;

Query Match 88.0%; Score 44; DB 2; Length 2126;
 Best Local Similarity 71.4%; Pred. No. 2.15e+00;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1222 PLEGGDI 1228
 |||||

QY 347 PLOGGSDV 353

RESULT 4

ID O61795 PRELIMINARY; PRT; 531 AA.
 AC 061795;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE R1E2.1 PROTEIN.
 GN R1E2.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE; 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA GOELA D., SCHEET P.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF067219; AAC17022.1; -;
 DR PFAM; PF00651; BTB; 1.
 DR PFAM; PF01344; Kelch; 4.
 DR PRINTS; PR00501; KELCHREPEAT.
 SQ SEQUENCE 531 AA; 58647 MW; 0F5C5F0A CRC32;

Query Match 86.0%; Score 43; DB 5; Length 531;
 Best Local Similarity 71.4%; Pred. No. 4.00e+00;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 301 PLOGGSDV 307
 QY 347 PLOGGSDV 353
 |||||

RESULT 5

ID O60506 PRELIMINARY; PRT; 623 AA.
 AC O60506;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE GRY-RBP.
 GN GRY-RBP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA DU G., YUAN J., PAN M., YAO H., CHEN J., QIANG B.;
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF037448; AAC12926.1; -;
 DR HSP; P19339; 2SXL.

US-09-376-430-2-24.rspt

Thu May 11 06:50:20 2000

```

DR PFAM: PF00005; ABC-tran; 1.
SQ SEQUENCE 790 AA; 87656 MW; 43608E61 CRC32;

Query Match
Best Local Similarity 86.0%; Score 43; DB 2; Length 790;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 93 PLOGGDV 99
QY 347 PLOGGDV 353

RESULT 8 PRELIMINARY; PRT: 2562 AA.
ID Q91957
AC Q91957; (T-EMBLrel. 01, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (T-EMBLrel. 08, Last annotation update)
DE XIN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
SEQUENCE FROM N.A.
RP TISSUE-CARDIAC MUSCLE;
RC WANG D.-Z., HU X., LIN J.L.-C., KITTEN G.T., SOLURSH M., LIN J.J.-C.;
RL Front. Biosci. 1:0-0(1996).
RN [2]
SEQUENCE FROM N.A.
RC TISSUE-CARDIAC MUSCLE;
RC WANG D.-Z., LIN J.J.-C.;
RA Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RL [3]
SEQUENCE OF 1762-2562 FROM N.A.
RP
RC TISSUE-LENS FIBERS; EGUCHI G.;
RA SAWADA K., AGATA K.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF051944; AAC06022.1;
DR EMBL: D82816; BAA11580.1;
KW Hypothetical protein.
SQ SEQUENCE 2562 AA; 284482 MW; E9FAD8B5 CRC32;

Query Match
Best Local Similarity 86.0%; Score 43; DB 13; Length 2562;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 211 PVOGGDV 217
QY 347 PLOGGDV 353

RESULT 9 PRELIMINARY; PRT: 8563 AA.
ID Q54297
AC Q54297; (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (T-EMBLrel. 01, Last annotation update)
DE POLYKETIDE SYNTHASE.
GN RAPA.
OS Streptomyces hygroscopicus.
OC Bacteria; Firmicutes; Actinobacteriales; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomyces.
RN [1]
SEQUENCE FROM N.A.
RP STRAIN-NRRL 5491.
RC MEDLINE: 9537237A
RX SCHWECHE T., APARICIO J.F., MOLNAR I., KOENIG A., KHAW L.E.,
RA SCHWECHE T., OLYNCK M., CAPREY P., CORTES J., LESTER J.B.,
RA HAYDOCK S.F., OLIVIERO J., LEADLEY P.F.,
RA BOEHM G.A., STAUNTON J.,
RT "The biosynthetic gene cluster for the polyketide immunosuppressant
rapamycin"; Acad. Sci. U.S.A. 92:7839-7843(1995).
RL Proc. Natl. Acad. Sci. U.S.A. 92:7839-7843(1995).

```

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NRRL 5491;
 RA MOLNAR I., APARICIO J.F., HAYDOCK S.F., EE KHAW L., SCHWECKE T.,
 RL KOENIG A., STAUNTON J., LEADLAY P.F., STAUNTON J., LEADLAY P.F.;
 RN Gene 0:0-0(0).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NRRL 5491;
 RA APARICIO J.F., MOLNAR I., SCHWECKE T., KOENIG A., HAYDOCK S.F.,
 RA EE KHAW L., STAUNTON J., LEADLAY P.F., LESTER J.B., BOEHM G.A.,
 RA STAUNTON J., LEADLAY P.F.;
 RL Gene 0:0-0(0).
 DR EMBL; X86780; CAA60460.1; -.
 DR HSSP; P08659; ILCL1.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 5.
 DR PROSITE; PS00606; B-KETOACYL SYNTHASE; 4.
 DR PROSITE; PS01162; OOR_ZETA-CRYSTAL; 2.
 DR PFAM; PF00698; AMP_BINDING; 1.
 DR PFAM; PF00107; adh_zinc; 4.
 DR PFAM; PF00501; AMP-binding; 1.
 DR PFAM; PF00550; pp-binding; 5.
 DR PFAM; PF00109; ketoacyl-synt; 4.
 DR PRINTS; PR00154; AMPBINDING.
 KW Transferase.
 SQ SEQUENCE 8563 AA; 899978 MW; 33DCD3BD CRC32;

Query Match 86.0%; Score 43; DB 2; Length 8563;
 Best Local Similarity 71.4%; Pred. No. 4.00e+00;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 2632 PLOGGEV 2638
 QY 347 PLOGGDV 353

RESULT 10
 ID Q54296 PRELIMINARY; PRT; 10223 AA.
 AC Q54296
 DT 01-NOV-1996 (TRENBLREL. 01, Created)
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
 DE POLYKETIDE SYNTHASE.
 GN RABP.
 OS Streptomyces hygroscopicus.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 RC Actinomycetales; Streptomycetaceae; Streptomyces.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NRRL 5491;
 RX MEDLINE; 95372374.
 RA SCHWECKE T., APARICIO J.F., MOLNAR I., KOENIG A., KHAW L.E.,
 RA HAYDOCK S.F., OLIVNYK M., CAFFEY P., CORTES J., LESTER J.B.,
 RA BOEHM G.A., STAUNTON J., LEADLAY P.F.;
 RT "The biosynthetic gene cluster for the polyketide immunosuppressant
 rapamycin".
 RL Proc. Natl. Acad. Sci. U.S.A. 92:7839-7843(1995).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NRRL 5491;
 RX MEDLINE; 96186895.
 RA MOLNAR I., APARICIO J.F., HAYDOCK S.F., EE KHAW L., SCHWECKE T.,
 RA KOENIG A., STAUNTON J., LEADLAY P.F.;
 RT "Organisation of the biosynthetic gene cluster for rapamycin in
 Streptomyces hygroscopicus: analysis of genes flanking the polyketide
 synthase".
 RL Gene 169:1-7(1996).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NRRL 5491;
 RX MEDLINE; 96186896.
 RA APARICIO J.F., MOLNAR I., SCHWECKE T., KOENIG A., HAYDOCK S.F.,

RA EE KHAW L., STAUNTON J., LEADLAY P., LESTER J.B., BOEHM G.A.,
 RA STAUNTON J., LEADLAY P.F.;
 RT "Organization of the biosynthetic gene cluster for rapamycin in
 Streptomyces hygroscopicus: analysis of the enzymatic domains in the
 modular polyketide synthase".
 RL Gene 169:9-16(1996).
 DR EMBL; X86780; CAA60459.1; -.
 DR HSSP; P19656; JAFH.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 6.
 DR PROSITE; PS00606; B-KETOACYL SYNTHASE; 6.
 DR PROSITE; PS01162; OOR_ZETA-CRYSTAL; 1.
 DR PFAM; PF00109; ketoacyl-synt; 6.
 DR PFAM; PF00698; Acyl-transf; 6.
 DR PFAM; PF00107; adh_zinc; 1.
 DR PFAM; PF00550; pp-binding; 6.
 KW Transferase.
 SQ SEQUENCE 10223 AA; 1072181 MW; 2C90396C CRC32;

Query Match 86.0%; Score 43; DB 2; Length 10223;
 Best Local Similarity 71.4%; Pred. No. 4.00e+00;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 4593 PLOGGEV 4599
 QY 347 PLOGGDV 353

RESULT 11
 ID O55715 PRELIMINARY; PRT; 165 AA.
 AC O55715
 DT 01-JUN-1998 (TRENBLREL. 06, Created)
 DT 01-JUN-1998 (TRENBLREL. 06, Last sequence update)
 DE HYPOTHETICAL 18.4 KD PROTEIN.
 OS Chilo iridescent virus (CIV) (Insect iridescent virus type 6).
 OC Viruses; dsDNA viruses, no RNA stage; Iridoviridae; Iridovirus.
 [1]
 RP SEQUENCE FROM N.A.
 RA BAHU U., TIDONA C.A., DARAI G.;
 RL Virus Genes 0:0-0(1997).
 DR EMBL; AF003534; AAB94426.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 165 AA; 18412 MW; 2C830341 CRC32;

Query Match 84.0%; Score 42; DB 14; Length 165;
 Best Local Similarity 71.4%; Pred. No. 7.37e+00;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 124 PLOGGEV 130
 QY 347 PLOGGDV 353

RESULT 12
 ID Q9YBK1 PRELIMINARY; PRT; 175 AA.
 AC Q9YBK1
 DT 01-NOV-1999 (TRENBLREL. 12, Created)
 DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)
 DE 175AA LONG HYPOTHETICAL PROTEIN.
 GN APE1597.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Aeropyrum.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K1
 RA KAWARABAYASHI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
 RA JIN N. K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
 RA HOSOIYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
 RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
 RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
 RA NOMURA N., SAKO Y., KIKUCHI H.;

US-09-376-430-2-24.ispt

Thu May 11 06:50:20 2000

RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res 6.83-101(1999).
 DR EMBL: AF000062; BAA80597.1; -. E8479BCB CRC32;
 SQ SEQUENCE 175 AA; 18953 MW; 18953 MW;

Query Match 84.0%; Score 42; DB 1; Length 175;
 Best Local Similarity 83.3%; Pred. No. 7.37e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 158 PLOGGE 163
 QY 347 PLOGGD 352

RESULT 13 PRELIMINARY; PRT; 321 AA.

ID Q9YGL4
 AC Q9YGL4
 DT 01-MAY-1999 (TRENBLrel. 10, Created)
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
 DE PALM PROTEIN.
 DE PALM.
 GN Gallus gallus (Chicken).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
 OC Neognathae; Galliformes; Phasianidae; Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 99030840.
 RA KUTZLEB C., SANDERS G., YAMAMOTO R., WANG X., LICHTER B.,
 RA PETRASCH-PARWEZ E., KILIMANN M.W.;
 RT "Paralemin, a prenyl-palmitoyl-anchored phosphoprotein abundant in
 RT neurons and implicated in plasma membrane dynamics and cell process
 RT formation.";
 RT J. Cell Biol. 143:795-813(1998).
 RL EMBL: Y14769; CAB37360.1; -.
 DR EMBL: Y14769; CAB37360.1; -.
 SQ SEQUENCE 321 AA; 34983 MW; 7CABA4870 CRC32;

Query Match 84.0%; Score 42; DB 13; Length 321;
 Best Local Similarity 71.4%; Pred. No. 7.37e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 282 PLOGDEV 288
 QY 347 PLOGGDV 353

RESULT 14 PRELIMINARY; PRT; 340 AA.

ID Q9XZX6
 AC Q9XZX6;
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
 DE L2743.9 PROTEIN.
 DE L2743.9.
 GN Leishmania major.
 OS Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MURPHY L., QUAIL M., LAWSON D., HARRIS D., RAJANDREAM M., IVENS A.,
 RA BARRELL B.;
 RA BARRELL B.; (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL031910; CAB42382.1; -.
 SQ SEQUENCE 340 AA; 36280 MW; CC8B931B CRC32;

Query Match 84.0%; Score 42; DB 5; Length 340;
 Best Local Similarity 83.3%; Pred. No. 7.37e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 323 PLOGGE 328
 QY 347 PLOGGD 352

RESULT 15 PRELIMINARY; PRT; 342 AA.
 ID Q9YGL5;
 AC Q9YGL5;
 DT 01-MAY-1999 (TRENBLrel. 10, Created)
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
 DE PALM PROTEIN.
 DE PALM.
 GN Gallus gallus (Chicken).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
 OC Neognathae; Galliformes; Phasianidae; Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 99030840.
 RA KUTZLEB C., SANDERS G., YAMAMOTO R., WANG X., LICHTER B.,
 RA PETRASCH-PARWEZ E., KILIMANN M.W.;
 RT "Paralemin, a prenyl-palmitoyl-anchored phosphoprotein abundant in
 RT neurons and implicated in plasma membrane dynamics and cell process
 RT formation.";
 RT J. Cell Biol. 143:795-813(1998).
 RL EMBL: Y14769; CAB37359.1; -.
 DR EMBL: Y14769; CAB37359.1; -.
 SQ SEQUENCE 342 AA; 37189 MW; 92BC2070 CRC32;

Query Match 84.0%; Score 42; DB 13; Length 342;
 Best Local Similarity 71.4%; Pred. No. 7.37e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 303 PLOGDEV 309
 QY 347 PLOGGDV 353

Search completed: Wed May 10 14:21:31 2000
 Job time : 243 secs.

US-09-376-430-2-25.rag

Thu May 11 06:50:21 2000

Infectious rubella vir 2.64e+02
 S. aureus D-alanine tr 3.50e+02
 Bacillus thuringiensis 3.50e+02
 E. coli colitose or gl 3.50e+02
 Tip adhesin protein. 3.50e+02
 A14 SLG protein. 3.50e+02
 Amino acid sequence of 3.50e+02
 Urochloa panicoides C4 3.50e+02
 Urochloa panicoides C4 3.50e+02
 Urochloa panicoides C4 3.50e+02
 Human beta galactosida 3.50e+02
 Streptococcus pneumonia 3.50e+02
 TGF-alpha-PE40-ab modi 4.63e+02
 TGF-alpha-PE40ab. 4.63e+02
 TGF-57-PE40 encoded by 4.63e+02
 Sequence encoded by al 4.63e+02
 Immunoprotein PEX46. 4.63e+02
 Pseudomonas exotoxin (4.63e+02
 Pseudomonas exotoxin (4.63e+02
 Pseudomonas sp. exotox 4.63e+02
 Pseudomonas sp. exotox 4.63e+02
 Human tumour progressi 4.63e+02

1 R79048
 2205 1 W73375
 282 1 W60218
 357 1 W88310
 374 1 R89329
 437 1 R41226
 437 1 Y13375
 585 1 R88552
 621 1 R88550
 624 1 R88551
 677 1 W70900
 763 1 W88458
 763 1 R06448
 420 1 W19870
 420 1 W19869
 420 1 R48950
 467 1 R04920
 549 1 R40108
 613 1 R40113
 613 1 W92912
 613 1 W92915
 613 1 W26667
 1533 1

24 41 75.9
 25 40 74.1
 26 40 74.1
 27 40 74.1
 28 40 74.1
 29 40 74.1
 30 40 74.1
 31 40 74.1
 32 40 74.1
 33 40 74.1
 34 40 74.1
 35 39 72.2
 36 39 72.2
 37 39 72.2
 38 39 72.2
 39 39 72.2
 40 39 72.2
 41 39 72.2
 42 39 72.2
 43 39 72.2
 44 39 72.2
 45 39 72.2

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd
 MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed May 10 14:30:26 2000; MasPar time 2.82 Seconds
 58.752 Million cell updates/sec

Tabular output not generated.
 Title: >US-09-376-430-2
 Description: (363-369) from US09376430A.pep (25 of 25)
 Perfect score: 54
 Sequence: 1 MNDRSIV 7

Scoring table: PAM 150
 Gap 11
 Searched: 188963 seqs, 23686106 residues
 Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-geneseq35
 1:geneseq
 Statistics: Mean 15.749; Variance 38.971; scale 0.404
 Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES
 Result No. Score Query Match Length DB ID Description Pred. No.

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	45	83.3	314	1	W65046	Thuja pinoresinol-lari	8.35e+01
2	44	81.5	301	1	R95589	Feline herpesvirus gly	1.12e+02
3	43	79.6	617	1	W94745	Mutant measles virus H	1.49e+02
4	43	79.6	617	1	R42391	Chicago 2 haemagglutinin	1.49e+02
5	43	79.6	617	1	R42390	Chicago 1 haemagglutinin	1.49e+02
6	43	79.6	617	1	W94752	Mutant measles virus H	1.49e+02
7	43	79.6	617	1	R42392	McI haemagglutinin	1.49e+02
8	43	79.6	617	1	R42389	San Diego haemagglutinin	1.49e+02
9	43	79.6	617	1	R42388	Consensus haemagglutinin	1.49e+02
10	43	79.6	617	1	R42387	H protein of attenuate	1.49e+02
11	43	79.6	617	1	R42386	Consensus haemagglutinin	1.49e+02
12	43	79.6	617	1	R42385	Consensus haemagglutinin	1.49e+02
13	43	79.6	617	1	R42384	JM haemagglutinin	1.49e+02
14	43	79.6	617	1	R42383	Moraten haemagglutinin	1.49e+02
15	43	79.6	617	1	R42382	Granulosis virus infec	1.49e+02
16	43	79.6	617	1	R42381	Protein kinase ch4	1.49e+02
17	43	79.6	617	1	R42380	Tuberos sclerosis 2 T	1.49e+02
18	43	79.6	617	1	R42379	Pyrolnitrin gene prot	1.49e+02
19	43	79.6	617	1	R42378	H. pylori secreted or	2.64e+02
20	43	79.6	617	1	R42377	H. pylori secreted or	2.64e+02
21	43	79.6	617	1	R42376	O antigen polymerase i	2.64e+02
22	43	79.6	617	1	R42375	Viral enhancing factor	2.64e+02
23	43	79.6	617	1	R42374	Transcription sequence	2.64e+02
24	43	79.6	617	1	R42373	Rubella virus RA27/3 N	2.64e+02

ALIGNMENTS

RESULT 1
 W65046 standard; Protein; 314 AA.

ID W65046;
 AC 28-SEP-1998 (first entry)
 DE Thuja pinoresinol-lariciresinol reductase PLR-Tp3.
 DT Pinoresinol-lariciresinol reductase; dirigent protein; lignan;
 KW transgenic plant; anticancer; cytostatic; antiviral; virucide;
 KW antibiotic; antioxidant; antifeedant.
 OS Thuja plicata.
 PN WO9820113-A1.
 PD 14-MAY-1998.
 PF 07-NOV-1997; U20391.
 PR 31-JUL-1997; US-054380.
 PR 08-NOV-1996; US-030522.
 PA (UNIV) UNIV WASHINGTON STATE RES FOUND.
 PI Davin LB, Dinkovakostova AT, Fujita M, Gang DR,
 PI Lewis NG, Sarkanen S;
 PI WPI: 98-286929/25.
 DR N-PSDB: V35184.
 DR New plant-derived dirigent proteins and pinoresinol-lariciresinol
 PT reductases - and related nucleic acid, vectors and transformants,
 PT used for stereospecific production of lignans, useful e.g. as
 PT anticancer or antiviral agents
 PS Claim 17: Page 128-129; 148pp: English.
 CC PLR-Tp3 comprises a (+)-pinoresinol/(+)-lariciresinol reductase of
 CC (P/LR) of Thuja plicata. P/LR enzymes catalyse the conversion of 4
 CC pinoresinol to lariciresinol and then to secoisolariciresinol. 4
 CC isoforms (see W65044-47) of T. plicata P/LR have been identified;
 CC the PLR-Tp1 isoform is (-)/(-)-P/LR. Further P/LR enzymes have
 CC been obtained from Forsythia intermedia (see W65038-43) and Tsuga
 CC heterophylla (see W65048-49). The isolation of cDNAs encoding P/LR
 CC (see V35175-80 and V35182-87) and dirigent proteins (see V35180-71)
 CC permits the development of an efficient expression system for these
 CC enzymes, provides useful tools for examining the developmental
 CC regulation of lignan biosynthesis and permits the isolation of a wide
 CC range of organisms, including plants, in order to modify lignan
 CC biosynthesis. Optically pure lignans may have e.g. anticancer,
 CC antiviral, antioxidant, antibiotic or antifeedant activity.
 CC Sequence 314 AA;
 SQ

Query Match 83.3%; Score 45; DB 1; Length 314;
 Best Local Similarity 71.4%; Pred. No. 8.35e+01;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 259 MKDKSVY 265
 1:1:111

QY 363 MNDRSYV 369

RESULT 2
 ID R96589 standard; Protein; 301 AA.
 AC R96589;
 DT 18-DEC-1996 (first entry)
 DE Feline herpesvirus glycoprotein G.
 KW gp G; unique short region; deletion; insertion; glycoprotein G;
 KW cat; vector; delivery; antigen.
 OS Feline herpesvirus.
 PN W09613575-A1.
 PD 09-MAY-1996.
 PF 26-OCT-1995; U13975.
 PR 26-OCT-1994; US-329883.
 PA (SYTR) SYNTRO CORP.
 PI Cochran MD, McDonnell MW;
 DR WPI; 96-239489/24.
 DT N-PSDB: T28191.
 DT Recombinant feline herpes virus attenuated by alteration of the gE
 PT gene - and related homology vectors, useful in vaccines, also as
 PT vectors for delivering antigens or therapeutic agents to mammals and
 PT birds
 PS Example 2; Page 107; 162pp; English.
 CC Feline herpesvirus (FHV) can be attenuated by altering its
 CC glycoprotein G (gp G) gene, so that it no longer produces
 CC functional gp G (i.e. the present sequence). The attenuated virus
 CC is useful in vaccines to protect cats against FHV (and opt. against
 CC other diseases if appropriate DNA is inserted), and as a vector
 CC for delivering vaccinating antigens or therapeutic agents (e.g.
 CC antiseptic mols., ribozymes, interferon inducers, hormones, etc.)
 CC lymphokines, etc.) to other mammals (including humans) or birds.
 SQ Sequence 301 AA;

Query Match 81.5%; Score 44; DB 1; Length 301;
 Best Local Similarity 71.4%; Pred. No. 1.12e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 275 MVDRAVY 281
 QY 363 MNDRSYV 369

RESULT 3
 ID W94745 standard; Protein; 617 AA.
 AC W94745;
 DT 20-APR-1999 (first entry)
 DE Measles; H protein; F protein; antigen; mutant; vaccine; detection;
 KW Measles; H protein; F protein; antigen; mutant; vaccine; detection;
 OS Measles virus.
 PN W09855627-A1.
 PD 10-DEC-1998.
 PF 04-JUN-1998; J02481.
 PR 04-JUN-1997; JP-184285.
 PA (OSAU) UNIV OSAKA.
 PI Kawanishi H, Ueda S, Watanabe M;
 DR WPI; 99-070272/06.
 DT N-PSDB; X05589.
 DT Mutant measles virus H and F protein antigens and genes encoding
 PT them - for production of attenuated virus or genetic vaccines and of
 PT diagnostic reagents
 PS Claim 1; Page 46-50; 93pp; Japanese.
 CC This represents a mutant measles virus H protein antigen. This protein
 CC is isolated from the CAM-70 (attenuated) strain of measles virus. The
 CC invention provides mutant measles virus antigenic proteins such as the
 CC virus and the F protein from the CAM-70 (attenuated) or NA (epidemic) strain of measles
 CC sequences encoding the proteins. These antigenic proteins are used in the
 CC production of vaccines for measles and the production of diagnostic
 CC reagents (e.g. measles antigens for the detection of specific antibodies
 CC in the blood of patients). The vaccines may be recombinant or other live

CC attenuated measles virus strains or may be genetic vaccines (e.g.
 CC adenovirus vectors). Specifically claimed peptide fragments (W94746-751)
 CC of this H protein can also be used for the production of vaccines and
 CC diagnostic reagents.
 SQ Sequence 617 AA;

Query Match 79.6%; Score 43; DB 1; Length 617;
 Best Local Similarity 71.4%; Pred. No. 1.49e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 31 MIDRPVY 37
 QY 363 MNDRSYV 369

RESULT 4
 ID R42391 standard; Protein; 617 AA.
 AC R42391;
 DT 13-MAY-1994 (first entry)
 DE Chicago 2 haemagglutinin.
 KW Haemagglutinin; HA; fusion glycoprotein; wild-type;
 KW Measles virus; vaccine; infection; consensus polypeptide.
 OS Measles virus; vaccine; infection; consensus polypeptide.
 PN W09321325-A.
 PD 28-OCT-1993.
 PF 08-APR-1993; U03209.
 PR 08-APR-1992; US-866033.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 PI Bellini WJ, Rota JS;
 DR WPI; 93-351735/44.
 DT N-PSDB; Q51092.
 DT Haemagglutinin and fusion glycoprotein of several wild-type
 PT measles strains - used to construct vaccines for measles
 PT infection
 PS Claim 10; Page 42-44; 119pp; English.
 CC HA (Q51088-94) and fusion glycoprotein (Q51095-97) sequences of
 CC several wild-type measles strains are given. Shared amino acid
 CC variations in wild-type measles glycoproteins are identified in five
 CC wild-type measles viruses. A consensus polypeptide, the amino acid
 CC sequence of which reflects variation common to more than one wild-
 CC type strain, is the basis for constructing live attenuated vaccines,
 CC or recombinant vaccines to replace older, less efficacious vaccines.
 CC Immunological reagents useful in differentiating wild-type measles
 CC strains from other known strains can also be produced.
 SQ Sequence 617 AA;

Query Match 79.6%; Score 43; DB 1; Length 617;
 Best Local Similarity 71.4%; Pred. No. 1.49e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 31 MIDRPVY 37
 QY 363 MNDRSYV 369

RESULT 5
 ID R42390 standard; Protein; 617 AA.
 AC R42390;
 DT 13-MAY-1994 (first entry)
 DE Chicago 1 haemagglutinin.
 KW Haemagglutinin; HA; fusion glycoprotein; wild-type;
 KW Measles virus; vaccine; infection; consensus polypeptide.
 OS Measles virus; vaccine; infection; consensus polypeptide.
 PN W09321325-A.
 PD 28-OCT-1993.
 PF 08-APR-1993; U03209.
 PR 08-APR-1992; US-866033.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 PI Bellini WJ, Rota JS;
 DR WPI; 93-351735/44.
 DT N-PSDB; Q51091.
 DT Haemagglutinin and fusion glycoprotein of several wild-type
 PT measles strains - used to construct vaccines for measles

US-09-376-430-2-25.rag

Thu May 11 06:50:21 2000

```

PS infection
PS Claim 9; Page 36-39; 119pp; English.
PS HA (Q51088-94) and fusion glycoprotein (Q51095-97) sequences of
CC several wild-type measles strains are given. Shared amino acid
CC variations in wild-type measles glycoproteins are identified in five
CC wild-type measles viruses. A consensus polypeptide, the amino acid
CC sequence of which reflects variation common to more than one wild-
CC type strain, is the basis for constructing live attenuated vaccines.
CC or recombinant vaccines to replace older, less efficacious vaccines.
CC immunological reagents useful in differentiating wild-type measles
CC strains from other known strains can also be produced.
SQ Sequence 617 AA;

Query Match 79.6%; Score 43; DB 1; Length 617;
Best Local Similarity 71.4%; Pred. No. 1.49e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 31 MIDRPYV 37
| | | | |
QY 363 MNDRSYV 369

RESULT 6 standard; Protein: 617 AA.
ID W94752 standard; Protein: 617 AA.
AC W94752; 1999 (first entry)
DT 20-APR-1999
DE Measles; H protein; F protein; antigen; mutant; vaccine; detection;
KW Measles; H protein; F protein; antigen; mutant; vaccine; detection;
KW diagnostic reagent; attenuated.
OS Measles virus.
PN W09855627-A1.
PD 10-DEC-1998.
PF 04-JUN-1998; JP-184285.
PR (OSAU) UNIV OSAKA.
PA Kawanishi H, Ueda S, Watanabe M;
PI WPI; 99-070272/06.
DR N-PSDB; X05590. Virus H and F protein antigens and genes encoding
PT mutant measles of attenuated virus or genetic vaccines and of
PT them - for production of attenuated virus or genetic vaccines and of
PT diagnostic reagents.
PT Claim 1; Page 60-63; 93pp; Japanese.
PS This represents a mutant measles virus H protein antigen. This protein
CC is isolated from the NA (epidemic) strain of measles proteins such as the
CC invention provides mutant measles virus antigenic proteins such as the
CC H protein from the CAM-70 (attenuated) or NA (epidemic) strain of measles
CC virus and the F protein from the CAM-70 or NA strain and nucleic acid
CC sequences encoding the proteins. These antigenic proteins are used in the
CC production of vaccines for measles and the production of diagnostic
CC reagents (e.g. measles antigens for the detection of specific antibodies
CC in the blood of patients). The vaccines may be recombinant or other live
CC attenuated measles virus strains or may be genetic vaccines (e.g.
CC adenovirus vectors). Specifically claimed peptide fragments (W94753-758)
CC of this H protein can also be used for the production of vaccines and
CC diagnostic reagents.
SQ Sequence 617 AA;

Query Match 79.6%; Score 43; DB 1; Length 617;
Best Local Similarity 71.4%; Pred. No. 1.49e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 31 MIDRPYV 37
| | | | |
QY 363 MNDRSYV 369

RESULT 7 standard; Protein: 617 AA.
ID R42392 standard; Protein: 617 AA.
AC R42392;
DT 13-MAY-1994 (first entry)
DE MCI haemagglutinin.
DE Haemagglutinin; HA; fusion glycoprotein; wild-type;
KW Measles virus; vaccine; infection; consensus polypeptide.
KW measles virus; vaccine; infection; consensus polypeptide.

OS Measles virus strain MCI.
PN W09321325-A.
PD 28-OCT-1993.
PF 08-APR-1993; U03209.
PR 08-APR-1992; US-866033.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Bellini WJ, Rota JS;
DR WPI; 93-351735/44.
DR N-PSDB; Q51093.
PT Haemagglutinin and fusion glycoprotein of several wild-type
CC measles strains - used to construct vaccines for measles
CC infection
CC Claim 1; Page 47-50; 119pp; English.
CC HA (Q51088-94) and fusion glycoprotein (Q51095-97) sequences of
CC several wild-type measles strains are given. Shared amino acid
CC variations in wild-type measles glycoproteins are identified in five
CC wild-type measles viruses. A consensus polypeptide, the amino acid
CC sequence of which reflects variation common to more than one wild-
CC type strain, is the basis for constructing live attenuated vaccines.
CC or recombinant vaccines to replace older, less efficacious vaccines.
CC immunological reagents useful in differentiating wild-type measles
CC strains from other known strains can also be produced.
SQ Sequence 617 AA;

Query Match 79.6%; Score 43; DB 1; Length 617;
Best Local Similarity 71.4%; Pred. No. 1.49e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 31 MIDRPYV 37
| | | | |
QY 363 MNDRSYV 369

RESULT 8 standard; Protein: 617 AA.
ID R42389 standard; Protein: 617 AA.
AC R42389;
DT 13-MAY-1994 (first entry)
DE San Diego haemagglutinin.
DE Haemagglutinin; HA; fusion glycoprotein; wild-type;
KW Measles virus; vaccine; infection; consensus polypeptide.
KW measles virus; vaccine; infection; consensus polypeptide.
OS W09321325-A.
PN W09321325-A.
PD 28-OCT-1993.
PF 08-APR-1993; U03209.
PR 08-APR-1992; US-866033.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Bellini WJ, Rota JS;
DR WPI; 93-351735/44.
DR N-PSDB; Q51090.
PT Haemagglutinin and fusion glycoprotein of several wild-type
CC measles strains - used to construct vaccines for measles
CC infection
CC Claim 2; Page 31-33; 119pp; English.
CC HA (Q51088-94) and fusion glycoprotein (Q51095-97) sequences of
CC several wild-type measles strains are given. Shared amino acid
CC variations in wild-type measles glycoproteins are identified in five
CC wild-type measles viruses. A consensus polypeptide, the amino acid
CC sequence of which reflects variation common to more than one wild-
CC type strain, is the basis for constructing live attenuated vaccines.
CC or recombinant vaccines to replace older, less efficacious vaccines.
CC immunological reagents useful in differentiating wild-type measles
CC strains from other known strains can also be produced.
SQ Sequence 617 AA;

Query Match 79.6%; Score 43; DB 1; Length 617;
Best Local Similarity 71.4%; Pred. No. 1.49e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 31 MIDRPYV 37
| | | | |
QY 363 MNDRSYV 369

```

RESULT 9
 ID R42388 standard; Protein; 617 AA.
 AC R42388 standard; Protein; 617 AA.
 DT 13-MAY-1994 (first entry)
 DE Consensus haemagglutinin sequence.
 KW Haemagglutinin; HA; fusion glycoprotein; wild-type;
 OS Measles virus; vaccine; infection; consensus polypeptide.
 PN W09321325-A.
 PD 28-OCT-1993.
 PF 08-APR-1993; U03209.
 PR 08-APR-1993; US-866033.
 PI (USSH) US DEPT HEALTH & HUMAN SERVICE.
 DR WPI: 93-351735/44.
 DT N-PSDB: Q51089.
 PT Haemagglutinin and fusion glycoprotein of several wild-type
 PT measles strains - used to construct vaccines for measles
 PS infection
 CC HA (Q51088-94) and fusion glycoprotein (Q51095-97) sequences of
 CC several wild-type measles strains are given. Shared amino acid
 CC variations in wild-type measles glycoproteins are identified in five
 CC wild-type measles viruses. A consensus polypeptide, the amino acid
 CC sequence of which reflects variation common to more than one wild-
 CC type strain, is the basis for constructing live attenuated vaccines,
 CC or recombinant vaccines to replace older, less efficacious vaccines.
 CC Immunological reagents useful in differentiating wild-type measles
 CC strains from other known strains can also be produced.
 SQ Sequence 617 AA;

Query Match 79.6%; Score 43; DB 1; Length 617;
 Best Local Similarity 71.4%; Pred. No. 1.49e-02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 31 MIDRPVY 37
 QY 363 MNDRSYV 369

RESULT 10
 ID R34541 standard; Protein; 617 AA.
 AC R34541 standard; Protein; 617 AA.
 DT 19-AUG-1993 (first entry).
 DE H protein of attenuated measles virus strain AIK-C.
 OS Paramyxoviridae; RNA virus; attenuation; vaccine.
 PN Human measles virus.
 PD 93-145503/18.
 PF 93-145503/18.
 PR N-PSDB: Q40480.
 PI Makino S, Mori T, Sasaki K;
 DR WPI: 93-145503/18.
 DT N-PSDB: Q40480.
 PT New attenuated measles vaccine virus strain - retains high
 PT immunogenicity with reduced pyrogenicity and having no
 PS neurological complications
 CC Disclosure: page 18-22; 47pp; English.
 CC The sequence of the entire genome of the attenuated measles
 CC virus strain AIK-C has been determined and contains 6 open reading
 CC frames. The sequence of the H protein was deduced from the fifth
 CC ORF. The virus retains high immunogenicity with reduced pyrogenicity
 CC and no neurological complications.
 CC See also R34537-R34540 and R39592.
 SQ Sequence 617 AA;

Query Match 79.6%; Score 43; DB 1; Length 617;
 Best Local Similarity 71.4%; Pred. No. 1.49e-02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 31 MIDRPVY 37
 QY 363 MNDRSYV 369

RESULT 11
 ID R49925 standard; Protein; 617 AA.
 AC R49925 standard; Protein; 617 AA.
 DT 13-MAY-1994 (first entry)
 DE Consensus haemagglutinin.
 KW Haemagglutinin; HA; fusion glycoprotein; wild-type;
 OS Measles virus; vaccine; infection; consensus polypeptide.
 PN Measles virus.
 PD 28-OCT-1993.
 PF 08-APR-1993; U03209.
 PR 08-APR-1993; US-866033.
 PI (USSH) US DEPT HEALTH & HUMAN SERVICE.
 DR WPI: 93-351735/44.
 DT N-PSDB: Q51089.
 PT Haemagglutinin and fusion glycoprotein of several wild-type
 PT measles strains - used to construct vaccines for measles
 PS infection
 CC HA (Q51088-94) and fusion glycoprotein (Q51095-97) sequences of
 CC several wild-type measles strains are given. Shared amino acid
 CC variations in wild-type measles glycoproteins are identified in five
 CC wild-type measles viruses. A consensus polypeptide, the amino acid
 CC sequence of which reflects variation common to more than one wild-
 CC type strain, is the basis for constructing live attenuated vaccines,
 CC or recombinant vaccines to replace older, less efficacious vaccines.
 CC Immunological reagents useful in differentiating wild-type measles
 CC strains from other known strains can also be produced.
 SQ Sequence 617 AA;

Query Match 79.6%; Score 43; DB 1; Length 617;
 Best Local Similarity 71.4%; Pred. No. 1.49e-02;

US-09-376-430-2-25.rag

Thu May 11 06:50:21 2000

```

PI Bellini WJ, Rota JS;
DR WPI; 93-351735/44. fusion glycoprotein of several wild-type
PT Haemagglutinin and fusion glycoprotein of several wild-type
PT measles strains - used to construct vaccines for measles
PT infection
PT Claim 1: Page 73-78; 119pp; English.
PS HA (Q51088-94) and fusion glycoprotein (Q51095-97) sequences of
CC several wild-type measles glycoproteins are given. Shared amino acid
CC variations in wild-type measles glycoproteins are identified in five
CC wild-type measles viruses. A consensus polypeptide, the amino acid
CC sequence of which reflects variation common to more than one wild-
CC type strain, is the basis for constructing live attenuated vaccines.
CC or recombinant vaccines to replace older, less efficacious vaccines.
CC immunological reagents useful in differentiating wild-type measles
CC strains from other known strains can also be produced.
CC Sequence 617 AA;
SQ
Query Match 79.6%; Score 43; DB 1; Length 617;
Best Local Similarity 71.4%; Pred. No. 1.49e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 31 MIDRPVY 37
QY 363 MNDRSYV 369

RESULT 12
ID R42393 standard; Protein; 617 AA.
AC R42393;
DE 13-MAY-1994 (first entry)
DE JM haemagglutinin.
DE Haemagglutinin; HA; fusion glycoprotein; wild-type.
KW measles virus; vaccine; infection; consensus polypeptide.
OS Measles virus strain JM.
PN WO9321325-A.
PD 28-OCT-1993.
PF 08-APR-1993; US-866033.
PR 08-APR-1992; US DEPT HEALTH & HUMAN SERVICE.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI Bellini WJ, Rota JS;
DR WPI; 93-351735/44.
DR N-PSDB; Q51094.
DR Haemagglutinin and fusion glycoprotein of several wild-type
PT measles strains - used to construct vaccines for measles
PT infection
PT Claim 12: Page 53-55; 119pp; English.
PS HA (Q51088-94) and fusion glycoprotein (Q51095-97) sequences of
CC several wild-type measles glycoproteins are given. Shared amino acid
CC variations in wild-type measles glycoproteins are identified in five
CC wild-type measles viruses. A consensus polypeptide, the amino acid
CC sequence of which reflects variation common to more than one wild-
CC type strain, is the basis for constructing live attenuated vaccines.
CC or recombinant vaccines to replace older, less efficacious vaccines.
CC immunological reagents useful in differentiating wild-type measles
CC strains from other known strains can also be produced.
CC Sequence 617 AA;
SQ
Query Match 79.6%; Score 43; DB 1; Length 617;
Best Local Similarity 71.4%; Pred. No. 1.49e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 31 MIDRPVY 37
QY 363 MNDRSYV 369

RESULT 13
ID R42387 standard; Protein; 617 AA.
AC R42387;
DE 13-MAY-1994 (first entry)
DE Moraten haemagglutinin.
DE Haemagglutinin; HA; fusion glycoprotein; wild-type;
KW measles virus; vaccine; infection; consensus polypeptide.

```

```

OS Measles virus strain Moraten.
PN WO9321325-A.
PD 28-OCT-1993.
PF 08-APR-1993; U03209.
PR 08-APR-1992; US-866033.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI Bellini WJ, Rota JS;
DR WPI; 93-351735/44.
DR N-PSDB; Q51088.
DR Haemagglutinin and fusion glycoprotein of several wild-type
PT measles strains - used to construct vaccines for measles
PT infection
PT Claim 1: Page 20-22; 119pp; English.
PS HA (Q51088-94) and fusion glycoprotein (Q51095-97) sequences of
CC several wild-type measles glycoproteins are given. Shared amino acid
CC variations in wild-type measles glycoproteins are identified in five
CC wild-type measles viruses. A consensus polypeptide, the amino acid
CC sequence of which reflects variation common to more than one wild-
CC type strain, is the basis for constructing live attenuated vaccines.
CC or recombinant vaccines to replace older, less efficacious vaccines.
CC immunological reagents useful in differentiating wild-type measles
CC strains from other known strains can also be produced.
CC Sequence 617 AA;
SQ
Query Match 79.6%; Score 43; DB 1; Length 617;
Best Local Similarity 71.4%; Pred. No. 1.49e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 31 MIDRPVY 37
QY 363 MNDRSYV 369

RESULT 14
ID W14777 standard; Protein; 898 AA.
AC W14777; 1997 (first entry)
DE Granulosis virus infectivity protein.
DE Infection; nuclear polyhedrosis virus; Granulosis virus genus.
OS Granulosis virus XCV alpha-4 strain.
FH Key Location/Qualifiers
FT cds 49.2745
FT /*tag= a
FT J09009972-A.
FT 14-JAN-1997.
PF 03-JUL-1995; JP-167481.
PR 03-JUL-1995; NUISANSO NOGYO KENKYU.
PA (NORQ ) NORINSUISANSO NOGYO KENKYU.
DR WPI; 97-126429/12.
DR N-PSDB; T14777.
DR Granulosis virus protein - enhances nuclear polyhedrosis virus
PT infectious activity
PS Claim 4; Page 10-13; 14pp; Japanese.
CC This sequence represents a protein which has a mol. wt. of ca. 100 KD and
CC which enhances the infectious activity of nuclear polyhedrosis virus.
CC This sequence is originated from a virus of Granulosis virus genus.
CC The protein enhances infectious activity of insecticidal NPV.
CC Sequence 898 AA;
SQ
Query Match 79.6%; Score 43; DB 1; Length 898;
Best Local Similarity 100.0%; Pred. No. 1.49e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 563 NDRSYV 568
QY 364 NDRSYV 369

RESULT 15
ID W98134 standard; Protein; 1163 AA.
AC W98134;
DE 21-JUN-1999 (first entry)
DE Protein kinase ch4.

```

US-09-376-430-2-25.rag

KW Elongation factor-2 kinase; eEF-2 kinase; protein kinase; ch4;
KW Inhibitor; therapy.
OS Mammalia.
PN WO990919-A2.
PD 25-FEB-1999.
PF 20-AUG-1998; U17272.
PR 20-AUG-1997; US-914999.
PA (UYN-) UNIV NEW JERSEY.
PI Hait WN, Pavur KS, Ryazanov AG;
DR WPI; 99-181050/15.
PT New isolated protein kinase, eEF-2 - used to develop agents for
PT controlling the amount, or activity of protein kinases, e.g. for
PT treating cancers or other hyperproliferative pathologies
PS Disclosure; Page 17-18; 195pp; English.
CC The present sequence is protein kinase ch4. Claimed isolated
CC nucleic acids of the invention code for a protein kinase which is a
CC member of a new superfamily characterised by: greater than 40%
CC sequence similarity with a eukaryotic elongation factor-2 (eEF-2)
CC kinase from any organism; and phosphorylation of an amino acid
CC within an alpha-helical domain of the target protein. The nucleic
CC acid preferably encodes eEF-2 kinase (see W98106-08), heart protein
CC kinase (see W98132), melanoma protein kinase (see W98133) or ch4
CC protein kinase. The protein kinases can be used in the search for
CC specific protein kinase inhibitors useful e.g. in cancer treatment.
SQ Sequence 1163 AA;

Query Match
Best Local Similarity 79.6%; Score 43; DB 1; Length 1163;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 349 VDERSYV 355
QY 363 MNDRSYV 369

Search completed: Wed May 10 14:30:33 2000
Job time : 7 secs.

US-09-376-430-2-25.rai

Thu May 11 06:50:21 2000

M P E R L H

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm
Wed May 10 14:30:51 2000; Maspar time 45.56 Seconds
1.991 Million cell updates/sec

Tabular output not generated.

Title: >US-09-376-430-2
(363-369) from US09376430A.pep (25 of 25)
Description: 54
Perfect score: 54
Sequence: 1 MNDRSYV 7

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1
Mean 14.906; Variance 39.090; scale 0.381

Statistics: Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	44	81.5	301	3	PCT-US95-1 Sequence 72, Applicati	6.97e+01
2	43	79.6	617	1	US-08-279- Sequence 4, Applicatio	9.23e+01
3	43	79.6	617	1	US-08-279- Sequence 6, Applicatio	9.23e+01
4	43	79.6	617	1	US-08-348- Sequence 8, Applicatio	9.23e+01
5	43	79.6	617	1	US-08-279- Sequence 14, Applicati	9.23e+01
6	43	79.6	617	1	US-08-279- Sequence 12, Applicati	9.23e+01
7	43	79.6	617	1	US-08-279- Sequence 21, Applicati	9.23e+01
8	43	79.6	617	1	US-08-505- Sequence 6, Applicatio	9.23e+01
9	43	79.6	617	2	US-08-279- Sequence 2, Applicati	9.23e+01
10	43	79.6	617	1	US-08-279- Sequence 10, Applicati	9.23e+01
11	43	79.6	617	1	US-08-854- Sequence 9, Applicatio	1.22e+02
12	42	77.8	211	1	US-08-379- Sequence 9, Applicatio	1.22e+02
13	42	77.8	211	1	US-08-428- Sequence 24, Applicati	1.61e+02
14	42	77.8	538	2	US-08-729- Sequence 50, Applicati	1.61e+02
15	41	75.9	53	2	US-08-469- Sequence 24, Applicati	1.61e+02
16	41	75.9	54	2	US-08-237- Sequence 50, Applicati	1.61e+02
17	41	75.9	54	2	US-08-456- Sequence 45, Applicati	1.61e+02
18	41	75.9	76	2	US-08-469- Sequence 6, Applicatio	1.61e+02
19	41	75.9	387	2	US-08-902- Sequence 2, Applicati	2.12e+02
20	41	75.9	2205	1	US-08-093- Sequence 81, Applicati	2.12e+02
21	41	75.9	370	2	US-08-846- Sequence 48, Applicati	2.12e+02
22	40	74.1	409	2	US-08-576- Sequence 48, Applicati	2.12e+02
23	40	74.1	409	2	US-08-576- Sequence 48, Applicati	2.12e+02

40 74.1 437 2 US-08-477- Sequence 9, Applicatio 2.12e+02
40 74.1 437 1 US-08-277- Sequence 6, Applicatio 2.12e+02
40 74.1 437 1 US-08-473- Sequence 6, Applicatio 2.12e+02
40 74.1 621 2 US-08-617- Sequence 2, Applicatio 2.12e+02
40 74.1 624 2 US-08-617- Sequence 2, Applicatio 2.12e+02
40 74.1 626 2 US-08-889- Sequence 2, Applicatio 2.12e+02
40 74.1 763 2 US-08-333- Sequence 9, Applicatio 2.78e+02
31 39 72.2 136 1 US-07-857- Sequence 77, Applicati 2.78e+02
31 39 72.2 387 1 US-07-626- Sequence 17, Applicati 2.78e+02
32 39 72.2 419 1 US-08-056- Sequence 4, Applicatio 2.78e+02
33 39 72.2 456 1 US-08-021- Sequence 6, Applicatio 2.78e+02
35 39 72.2 456 1 US-08-082- Sequence 6, Applicatio 2.78e+02
36 39 72.2 460 1 US-08-351- Sequence 8, Applicatio 2.78e+02
37 39 72.2 472 1 US-08-021- Sequence 10, Applicati 2.78e+02
38 39 72.2 508 1 US-08-021- Sequence 2, Applicatio 2.78e+02
39 39 72.2 538 1 US-08-457- Sequence 2, Applicatio 2.78e+02
40 39 72.2 538 1 US-08-463- Sequence 3, Applicatio 2.78e+02
41 39 72.2 599 1 US-08-463- Sequence 1, Applicatio 2.78e+02
42 39 72.2 613 2 US-08-235- Sequence 16, Applicati 2.78e+02
43 39 72.2 637 1 US-08-623- Sequence 7, Applicatio 2.78e+02
44 39 72.2 1497 1
45

ALIGNMENTS

RESULT 1 STANDARD; PRT; 301 AA.
ID PCT-US95-13975-72
XX xxxxxx
AC
XX
XX
DT
XX
DE
XX
XX
Sequence 72, Application PC/TUS9513975
Sequence 72, Application PC/TUS9513975
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
APPLICANT: McDonnell, Michael W.
TITLE OF INVENTION: Recombinant Feline Herpes virus
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM 330 466 DX2
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: PCT/US95/13975
FILING DATE: 26-OCT-1995
CLASSIFICATION:
PRIOR APPLICATION NUMBER: 08/329,883
FILING DATE: 26-OCT-1994
CLASSIFICATION INFORMATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P. 28,678
REGISTRATION NUMBER: 39118-PCT
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)391-0525
TELEFAX: (212)391-0525
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 301 amino acids
TYPE: amino acid
TOPOLOGY: linear

US-09-376-430-2-25.ra1

Page 2

Sequence 5, Application US/08279700
Sequence 6, Application US/08279700
Patent No. 5578448

MFP: 22313-0299
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER:

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
TELEPHONE/DOCKET NUMBER: 40399/140 NIHD
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 617 amino acids
TYPE: amino acid
TOPOLOGI: linear
MOLECULE TYPE: protein
SEQ SEQUENCE 617 AA; 69154 MW; 2111815 CN;
Query Watch

0; Gaps 0;

RESULT 4*
ID US-08-348-891A-6
XX
AC XXXXXX
STANDARD;
PRT; 617 AA.

Sequence 6, Application US/08348891A
Sequence 6, Application US/08348891A

US-09-376-430-2-25.rai

Thu May 11 06:50:21 2000

CC Patent No. 5654136
 CC GENERAL INFORMATION:
 CC APPLICANT: SASAKI, Keiko
 CC APPLICANT: MORI, Takayuki
 CC APPLICANT: MARINO, Satoshi
 CC TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE, SEQUENCE AND A METHOD FOR
 CC TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND ITS ABSOLUTE IDENTIFICATION
 CC TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION
 CC NUMBER OF SEQUENCES: 19
 CC NUMBER OF SEQUENCES: 19
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: YOUNG & THOMPSON
 CC STREET: 745 South 23rd Street
 CC CITY: Arlington
 CC STATE: Virginia
 CC COUNTRY: USA
 CC ZIP: 22202
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patent in Release #1.0, Version #1.30
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/348,891A
 CC FILING DATE: 25-NOV-1994
 CC CLASSIFICATION: 435
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 07/848,400
 CC FILING DATE: 10-MAR-1992
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: JP 3-293625
 CC FILING DATE: 14-OCT-1991
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: PAICH, Andrew J.
 CC REGISTRATION NUMBER: 32,925
 CC REFERENCE/DOCKET NUMBER: KP-7501
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 703-521-2297
 CC TELEFAX: 703-685-0573
 CC TELEX: 248425 EMBON
 CC INFORMATION FOR SEQ ID NO: 6:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 617 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC SEQUENCE 617 AA: 69340 MW: 2115500 CN;
 CC
 CC Query Match 79.6% Score 43; DB 1; Length 617;
 CC Best Local Similarity 71.4% Pred. No. 9.23e+01;
 CC Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC Db 31 MIDRPVY 37
 CC QY 363 MNDRSYV 369
 CC
 CC RESULT 5
 CC ID US-08-279-700-8 STANDARD; PRT: 617 AA.
 CC XX
 CC AC xxxxxx
 CC DT
 CC DE
 CC DX
 CC Sequence 8, Application US/08279700
 CC Patent No. 5578448
 CC GENERAL INFORMATION:
 CC APPLICANT: ROTA, Jennifer S.
 CC APPLICANT: BELLINI, WILLIAM J.
 CC TITLE OF INVENTION: MEASLES VIRUS GLYCOPROTEINS:
 CC TITLE OF INVENTION: VACCINE AND DETECTION METHOD THEREFOR
 CC NUMBER OF SEQUENCES: 22
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: FOLEY & LARDNER
 CC STREET: 1800 Diagonal Road, Suite 500
 CC CITY: Alexandria
 CC STATE: Virginia
 CC COUNTRY: USA
 CC ZIP: 22313-0299
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: FOLEY & LARDNER
 CC STREET: 1800 Diagonal Road, Suite 500
 CC CITY: Alexandria
 CC STATE: Virginia
 CC COUNTRY: USA
 CC ZIP: 22313-0299
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patent in Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/279,700
 CC FILING DATE: 435
 CC CLASSIFICATION: 435
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US/07/866,033A
 CC FILING DATE: 19920408
 CC CLASSIFICATION: 435
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: BENT, Stephen A.
 CC REGISTRATION NUMBER: 29,768
 CC REFERENCE/DOCKET NUMBER: 40399/140 NIHD
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (703)836-9300
 CC TELEFAX: (703)683-4109
 CC TELEX: 899149
 CC INFORMATION FOR SEQ ID NO: 8:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 617 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC SEQUENCE 617 AA: 69124 MW: 2107447 CN;
 CC
 CC Query Match 79.6% Score 43; DB 1; Length 617;
 CC Best Local Similarity 71.4% Pred. No. 9.23e+01;
 CC Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC Db 31 MIDRPVY 37
 CC QY 363 MNDRSYV 369
 CC
 CC RESULT 6
 CC ID US-08-279-700-14 STANDARD; PRT: 617 AA.
 CC XX
 CC AC xxxxxx
 CC DT
 CC DE
 CC DX
 CC Sequence 14, Application US/08279700
 CC Patent No. 5578448
 CC GENERAL INFORMATION:
 CC APPLICANT: ROTA, Jennifer S.
 CC APPLICANT: BELLINI, WILLIAM J.
 CC TITLE OF INVENTION: MEASLES VIRUS GLYCOPROTEINS:
 CC TITLE OF INVENTION: VACCINE AND DETECTION METHOD THEREFOR
 CC NUMBER OF SEQUENCES: 22
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: FOLEY & LARDNER
 CC STREET: 1800 Diagonal Road, Suite 500
 CC CITY: Alexandria
 CC STATE: Virginia
 CC COUNTRY: USA
 CC ZIP: 22313-0299
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS

US-09-376-430-2-25.ra1

Thu May 11 06:50:21 2000

```

MOLECULE TYPE: protein
ORIGINAL SOURCE: consensus HA polypeptide
STRAIN: consensus HA polypeptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "Xaa denotes Gln or His"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 19
OTHER INFORMATION: /note= "Xaa denotes Lys or Arg"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 176
OTHER INFORMATION: /note= "Xaa denotes Thr, Val or
OTHER INFORMATION: Ala"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 235
OTHER INFORMATION: /note= "Xaa denotes Glu or Gly"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 295
OTHER INFORMATION: /note= "Xaa denotes Lys or Arg"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 303
OTHER INFORMATION: /note= "Xaa denotes Glu or Gly"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 305
OTHER INFORMATION: /note= "Xaa represents Ser or Phe"
OTHER INFORMATION: /note= "Xaa denotes Ile or Val"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 306
OTHER INFORMATION: /note= "Xaa denotes Ile or Val"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 308
OTHER INFORMATION: /note= "Xaa denotes Ile or Val"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 320
OTHER INFORMATION: /note= "Xaa denotes Gln or Arg"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 339
OTHER INFORMATION: /note= "Xaa denotes Leu or Phe"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 348
OTHER INFORMATION: /note= "Xaa denotes Arg or Lys"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 367
OTHER INFORMATION: /note= "Xaa denotes Val or Ile"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 389
OTHER INFORMATION: /note= "Xaa denotes Lys or Arg"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 390
OTHER INFORMATION: /note= "Xaa denotes Ile or Asn"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 446
OTHER INFORMATION: /note= "Xaa denotes Ser or Thr"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 451
OTHER INFORMATION: /note= "Xaa denotes Val or Glu"
FEATURE:

```

CC	NAME/KEY: Modified-site	
CC	LOCATION: 485	↑
CC	OTHER INFORMATION: /note= "Xaa denotes Val or Ile"	
CC	FEATURE:	
CC	NAME/KEY: Modified-site	
CC	LOCATION: 501	
CC	OTHER INFORMATION: /note= "Xaa denotes Pro or Ser"	
CC	FEATURE:	
CC	NAME/KEY: Modified-site	
CC	LOCATION: 544	
CC	OTHER INFORMATION: /note= "Xaa denotes Ser or Asn"	
CC	FEATURE:	
CC	NAME/KEY: Modified-site	
CC	LOCATION: 546	
CC	OTHER INFORMATION: /note= "Xaa denotes Ser or Gly"	
CC	FEATURE:	
CC	NAME/KEY: Modified-site	
CC	LOCATION: 559	
CC	OTHER INFORMATION: /note= "Xaa denotes Ile or Val"	
CC	FEATURE:	
CC	NAME/KEY: Modified-site	
CC	LOCATION: 560	
CC	OTHER INFORMATION: /note= "Xaa denotes Lys or Arg"	
CC	FEATURE:	
CC	NAME/KEY: Modified-site	
CC	LOCATION: 562	
CC	OTHER INFORMATION: /note= "Xaa denotes Val, Ile or	
CC	OTHER INFORMATION: Phe"	
CC	FEATURE:	
CC	NAME/KEY: Modified-site	
CC	LOCATION: 593	
CC	OTHER INFORMATION: /note= "Xaa denotes His or Tyr"	
CC	FEATURE:	
CC	NAME/KEY: Modified-site	
CC	LOCATION: 616	
CC	OTHER INFORMATION: /note= "Xaa denotes Arg or Ser"	
CC	SEQUENCE 617 AA; 69110 MW; 2217966 CN;	
SQ	Query Match	79.6%; Score 43; DB 1; Length 617
	Best Local Similarity	71.4%; Pred. No. 9.23e+01;
	Matches	1; Mismatches 1; Indels
		5; Conservative
Db	31 MIDRPV 37	
	111111	
Qy	363 MNDASYV 369	
RESULT 9	STANDARD:	PRT; 617 AA.
ID	US-08-905-817-6	
XX	xxxxxx	
AC		
XX		
DI		
DE	Sequence 6, Application US/08905817	
XX	Sequence 6, Application US/08905817	
CC	Patent No. 5924777	
CC	GENERAL INFORMATION:	
CC	APPLICANT: SASAKI, Keiko	
CC	APPLICANT: MORI, Takayuki	
CC	APPLICANT: MAKINO, Satoshi	
CC	TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACC	
CC	TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTID	
CC	TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION	
CC	NUMBER OF SEQUENCES: 19	
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE: YOUNG & THOMPSON	
CC	STREET: 745 South 23rd Street	
CC	City: Arlington	
CC	STATE: Virginia	
CC	COUNTRY: USA	
CC	ZIP: 22202	

Registration No. 08905817

Sequence 0, pp. 1-17

Sequence 6, Application

PATENT NO. 3,024,000
GENERAL INFORMATION:

APPLICANT: SASAKI
APPLICANT: MORI

APPLICANT:
APPLICANT:

TITLE OF INVENTION

TITLE OF INVENTION

NUMBER OF SEQUENCES

CORRESPONDENCE: YOUNG

STREET: 745 So

CITY: Arlington
STATE: Virginia

COUNTRY: USA

ZIP: 77777

1

RESULT 10
 ID US-08-279-700-2
 XX STANDARD; PRT; 617 AA.
 AC xxxxxx
 XX
 DT
 XX
 DE Sequence 2, Application US/08279700
 XX
 CC Sequence 2, Application US/08279700
 CC Patent No. 5578448
 CC GENERAL INFORMATION:
 CC APPLICANT: ROTA, Jennifer S.
 CC APPLICANT: BELLINI, William J.
 CC TITLE OF INVENTION: WILD-TYPE MEASLES VIRUS GLYCOPROTEINS:
 CC TITLE OF INVENTION: VACCINE AND DETECTION METHOD THEREFOR
 CC NUMBER OF SEQUENCES: 22
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Foley & Lardner
 CC STREET: 1800 Diagonal Road, Suite 500
 CC CITY: Alexandria
 CC STATE: Virginia
 CC COUNTRY: USA
 CC ZIP: 22313-0299
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patent in Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:

US-09-376-430-2-25.ra1

Thu May 11 06:50:21 2000

QY 363 MNDRSYV 369

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)836-9300
 TELEFAX: (703)683-4109
 TELEX: 899149
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 617 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE 617 AA: 69385 MW: 2101975 CN:

Query Match 79.6%; Score 43; DB 1; Length 617;
 Best Local Similarity 71.4%; Pred. No. 9.23e+01; Indels 0; Gaps 0;
 Matches 5; Conservative 1; Mismatches 1;

Db 31 MIDRSYV 37
 QY 363 MNDRSYV 369

RESULT 12 US-08-854-029-9 STANDARD; PRT; 211 AA.

ID XX
 AC AC
 DT XX
 DX XX
 DE XX
 CC CC

Sequence 9, Application US/08854029

Sequence 9, Application US/08854029

Patent No. 594074
 GENERAL INFORMATION:
 APPLICANT: Cold Spring Harbor Laboratory
 TITLE OF INVENTION: No. 594074el Human cdc25 Genes, Encoded
 TITLE OF INVENTION: Products and Uses Thereof
 NUMBER OF SEQUENCES: 31
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FOLEY, HOAG & ELIOT LLP
 STREET: One Post Office Square
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: U.S.A.

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII(text)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/854.029
 FILING DATE: 2 MAY 1997

CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Vincent, Matthew P.
 REGISTRATION NUMBER: 36,709
 REFERENCE/DOCKET NUMBER: MIV-019.08
 TELEPHONE: 617-832-1242
 TELEFAX: 617-832-7000

INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 211 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 SEQUENCE 211 AA: 24718 MW: 249937 CN;

Query Match 77.8%; Score 42; DB 2; Length 211;
 Best Local Similarity 85.7%; Pred. No. 1.22e+02; Indels 0; Gaps 0;
 Matches 6; Conservative 0; Mismatches 1;

Db 170 MNDRSYV 176

RESULT 13 US-08-379-685-9 STANDARD; PRT; 211 AA.

ID XX
 AC AC
 DT XX
 DX XX
 DE XX
 CC CC

Sequence 9, Application US/08379685
 Sequence 9, Application US/08379685

Patent No. 5770423
 GENERAL INFORMATION:
 APPLICANT: Cold Spring Harbor Laboratory
 TITLE OF INVENTION: No. 5770423el Human cdc25 Genes, Encoded
 TITLE OF INVENTION: Products and Uses Thereof
 NUMBER OF SEQUENCES: 31
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD
 STREET: 60 State Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: U.S.A.

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII(text)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/379.685
 FILING DATE: 26 January 1995

CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Vincent, Matthew P.
 REGISTRATION NUMBER: 36,709
 REFERENCE/DOCKET NUMBER: MII-019-DV
 TELEPHONE: 617-227-7400
 TELEFAX: 617-227-5941

INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 211 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 SEQUENCE 211 AA: 24718 MW: 249937 CN;

Query Match 77.8%; Score 42; DB 1; Length 211;
 Best Local Similarity 85.7%; Pred. No. 1.22e+02; Indels 0; Gaps 0;
 Matches 6; Conservative 0; Mismatches 1;

SQ SEQUENCE 211 AA: 24718 MW: 249937 CN;

Db 170 MNDRSYV 176
 QY 363 MNDRSYV 369

RESULT 14 US-08-428-415-9 STANDARD; PRT; 211 AA.

ID XX
 AC AC
 DT XX
 DX XX
 DE XX
 CC CC

Sequence 9, Application US/08428415
 Sequence 9, Application US/08428415

Patent No. 5756335
 GENERAL INFORMATION:
 APPLICANT: Cold Spring Harbor Laboratory

Thu May 11 06:50:21 2000

US-09-376-430-2-25.ra1

TITLE OF INVENTION: No. 5756335el Human cdc25 Genes, Encoded
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428.415
FILING DATE: 24 April 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MII-019CP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-7400
TELEFAX: 617-227-5941

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 211 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE 211 AA: 24718 MW; 249937 CN;

Query Match
Best Local Similarity 77.8%; Score 42; DB 1; Length 211;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 170 MNDRSV 176
QY 363 MNDRSV 369

RESULT 15
ID US-08-729-214-24 STANDARD; PRT: 538 AA.

Sequence 24, Application US/08729214
Sequence 24, Application US/08729214
Patent No. 5817502
GENERAL INFORMATION:
APPLICANT: Ligon, James M.
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Hammer, Phillip E.
APPLICANT: van Pee, Karl-Heinz
TITLE OF INVENTION: Genes for the synthesis of
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 520 White Plains Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/729,214
FILING DATE: TBA
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC 1506/CIPS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 538 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 538 AA: 60908 MW; 1433064 CN;

Query Match
Best Local Similarity 77.8%; Score 42; DB 2; Length 538;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 192 LNDRGYI 198
QY 363 MNDRSV 369

Search completed: Wed May 10 14:31:45 2000
Job time : 54 secs.

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/729,214
FILING DATE: TBA
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC 1506/CIPS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 538 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 538 AA: 60908 MW; 1433064 CN;

Query Match
Best Local Similarity 77.8%; Score 42; DB 2; Length 538;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 192 LNDRGYI 198
QY 363 MNDRSV 369

Search completed: Wed May 10 14:31:45 2000
Job time : 54 secs.

US-09-376-430-2-25.rpt

Thu May 11 06:50:22 2000

OY 363 MNDRSYV 369

Query Match 85.2%; Score 46; DB 1; Length 2638; Indels 0; Gaps 0;
Best Local Similarity 57.1%; Pred. No. 6.84e+00; Mismatches 0;
Matches 3; Conservative 0;

Db 1800 MEERGTV 1806
I:::I
OY 363 MNDRSYV 369

RESULT 5
ENTRY
TITLE
ORGANISM
DATE
G70021 #type complete
hypothetical protein yusn - Bacillus subtilis
#formal_name Bacillus subtilis
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998

ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:92087490
#contents
#note
#GENETICS
#gene
SUMMARY
Query Match 83.3%; Score 45; DB 2; Length 1123; Indels 0; Gaps 0;
Best Local Similarity 71.4%; Pred. No. 1.16e+01; Mismatches 1;
Matches 5; Conservative 1;

#journal
#title
#cross-references MUID:98044033
#accession
#status

#molecule_type DNA
#residues
#cross-references GB:299120; GB:AL009126; NID:G2635613; PID:ell84364;
#experimental_source strain
#length 110 #molecular-weight 13138 #checksum 430
#gene
#summary

Query Match 83.3%; Score 45; DB 2; Length 110; Indels 0; Gaps 0;
Best Local Similarity 71.4%; Pred. No. 1.16e+01; Mismatches 1;
Matches 5; Conservative 1;

Db 21 MNDREDFV 27

OY 363 MNDRSYV 369

RESULT 6
ENTRY
TITLE
ORGANISM
DATE
D36790 #type complete
hypothetical protein ORF39 - ictaluriid herpesvirus 1 (strain
#formal_name ictaluriid herpesvirus 1
#host ictalurus punctatus (Channel catfish)
17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change
09-Sep-1997

ACCESSIONS
REFERENCE
#authors
#submission
#description
#accession
#molecule_type DNA
#residues
#cross-references GB:M75136; NID:G331209; PID:G331249
D36790 A39447
Davison, A.J.
submitted to GenBank, January 1992
Channel catfish virus: a new type of herpesvirus.

39 #length 1123 #molecular-weight 123008 #checksum 4926
#GENETICS
#gene
SUMMARY
Query Match 83.3%; Score 45; DB 2; Length 1123; Indels 0; Gaps 0;
Best Local Similarity 83.3%; Pred. No. 1.16e+01; Mismatches 0;
Matches 5; Conservative 1;

Db 534 MNDRAY 539
I:::I
OY 363 MNDRSYV 368

RESULT 7
ENTRY
TITLE
ORGANISM
DATE
H69906 #type complete
macrolide glycosyltransferase homolog yojK - Bacillus
#formal_name Bacillus subtilis
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
22-Jun-1999
H69906
ACCESSIONS
REFERENCE
#authors

#molecule_type DNA
#residues
#cross-references GB:299120; GB:AL009126; NID:G2635613; PID:ell84364;
#experimental_source strain
#length 110 #molecular-weight 13138 #checksum 430
#gene
#summary

Query Match 83.3%; Score 45; DB 2; Length 110; Indels 0; Gaps 0;
Best Local Similarity 71.4%; Pred. No. 1.16e+01; Mismatches 1;
Matches 5; Conservative 1;

Db 21 MNDREDFV 27

Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
 Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
 Sekowska, A.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
 B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Shin, B.S.; Soldo,
 Takamaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, H.;
 Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.;
 Vandenberg, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
 Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
 Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
 K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
 Yoshioka, H.; Danchin, A.
 Nature (1997) 390:249-256
 The complete genome sequence of the Gram-positive bacterium
 Bacillus subtilis.
 Nature (1997) 390:249-256
 #cross-references EMBL:98044033
 #status Preliminary; nucleic acid sequence not shown;
 translation not shown
 #molecule_type DNA
 #residues 1-287 #label KUN
 #cross-references GB:299114; GB:AL009126; NID:g2634230;
 PID:CAR13834.1; PID:el185414; PID:g2634335
 #experimental_source strain 168
 #gene yojk
 #superfamily glycosyltransferase
 #length 287 #molecular-weight 31990 #checksum 4376
 Query Match 81.5%; Score 44; DB 2; Length 287;
 Best Local Similarity 83.3%; Pred. No. 1.94e+01;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 245 MNRYSY 250
 QY 363 MNRDSY 368

RESULT 8
 ENTRY WD-repeat protein T419.10 - Arabidopsis thaliana
 TITLE #formal_name Arabidopsis thaliana #common_name mouse-ear
 ORGANISM Arabidopsis thaliana
 DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
 ACCESSION T01391
 REFERENCE T419.10
 #authors Parnell, L.D.; Gnoj, L.; de la Bastide, M.; Hameed, A.;
 Habermann, K.; Schutz, K.; Huang, E.; Gottesman, T.;
 Bedhia, N.N.; McCombie, W.R.
 Submitted to the EMBL Data Library, May 1998
 Genomic sequence of BAC T419 from Arabidopsis thaliana,
 chromosome IV, near 10.6 cm.
 #status Translated from GB/EMBL/DDBJ
 #molecule_type DNA
 #residues 1-555 #label PAR
 #cross-references EMBL:AF069442; NID:g3242970; PID:g3924603
 #experimental_source cultivar Columbia
 #map_position 4
 #introns 4
 #note 68/2; 208/3; 248/3; 274/3; 333/3; 359/3; 419/2; 472/2; 492/3
 SUMMARY T419.10
 #length 555 #molecular-weight 62800 #checksum 9446
 Query Match 81.5%; Score 44; DB 2; Length 555;
 Best Local Similarity 57.1%; Pred. No. 1.94e+01;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 187 MDSRAY 193
 QY 363 MNRDSY 369

RESULT 9
 ENTRY rho-type GTPase-activating protein DBM1 - yeast
 TITLE #type complete
 ORGANISM Saccharomyces cerevisiae
 DATE 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change
 ACCESSION S48535
 REFERENCE S48535
 #authors Chen, G.; Zheng, L.; Chan, C.S.M.
 Submitted to the EMBL Data Library, March 1994
 Molecular characterization of Dbm1, a novel rhoGAP protein in
 yeast.
 #accession S48535
 #molecule_type DNA
 #residues 1-1007 #label CHE
 #cross-references EMBL:U07421; NID:g460972; PID:g460973
 #authors Wiemann, S.; Rechmann, S.; Benes, V.; Voss, H.; Schwager, C.;
 Vleck, C.; Stegemann, J.; Zimmermann, J.; Erfle, H.; Paces,
 V.; Ansoerge, W.; Zimmermann, J.; Erfle, H.; Paces,
 Submitted to the EMBL Data Library, August 1995
 Sequencing of 51 kilobases on the right arm of
 from S. cerevisiae reveals 30 open reading frames.
 #accession S60987
 #molecule_type DNA
 #residues 1-1007 #label WIE
 #cross-references EMBL:X90518; NID:g1050808; PID:g1050813
 #authors Benes, V.; Andrade, M.A.; Rechmann, S.; Teodoru, C.; Banreivi,
 A.; Sander, C.; Valencia, A.; Ansoerge, W.; Voss, H.
 Submitted to the EMBL Data Library, December 1995
 Nucleotide sequence and analysis of a 130 kb fragment of
 yeast chromosome XV.
 #accession S61683
 #molecule_type DNA
 #residues 1-1007 #label BEN
 #cross-references EMBL:X94335; NID:g1262139; PID:e217748; PID:g1164971
 #authors Voss, H.; Benes, V.; Rechmann, S.; Teodoru, C.; Schwager, C.;
 Paces, V.; Ansoerge, W.
 Submitted to the Protein Sequence Database, July 1996
 #accession S67012
 #molecule_type DNA
 #residues 1-1007 #label VOS
 #cross-references EMBL:275035; MIPS:YOR127w; NID:g1420329; PID:e252363;
 #experimental_source strain S288C
 #accession S63860
 #authors Wiemann, S.; Rechmann, S.; Benes, V.; Voss, H.; Schwager, C.;
 Vleck, C.; Stegemann, J.; Zimmermann, J.; Erfle, H.; Paces,
 V.; Ansoerge, W.
 Submitted to the EMBL Data Library, August 1995
 Sequencing and analysis of 51 kb on the right arm of
 chromosome XV from Saccharomyces cerevisiae reveals 30 open
 reading frames.
 #cross-references MIMD:97060020
 #accession S63864
 #status Nucleic acid sequence not shown; translation not shown
 #molecule_type DNA
 #residues 1-1007 #label WIV
 #cross-references EMBL:X90518; NID:g1050808; PID:g1050813
 #note the nucleotide sequence was submitted to the EMBL Data
 Library, August 1995
 #accession JC6097
 #authors Chen, G.; Zheng, L.; Chan, C.S.M.
 Submitted to the EMBL Data Library, August 1995
 The LIM domain-containing dbm1 GTPase-activating protein is
 required for normal cellular morphogenesis in Saccharomyces
 cerevisiae.
 #cross-references MIMD:96239492

RESULT 9
 ENTRY rho-type GTPase-activating protein DBM1 - yeast
 TITLE #type complete
 ORGANISM Saccharomyces cerevisiae
 DATE 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change
 ACCESSION S48535
 REFERENCE S48535
 #authors Chen, G.; Zheng, L.; Chan, C.S.M.
 Submitted to the EMBL Data Library, March 1994
 Molecular characterization of Dbm1, a novel rhoGAP protein in
 yeast.
 #accession S48535
 #molecule_type DNA
 #residues 1-1007 #label CHE
 #cross-references EMBL:U07421; NID:g460972; PID:g460973
 #authors Wiemann, S.; Rechmann, S.; Benes, V.; Voss, H.; Schwager, C.;
 Vleck, C.; Stegemann, J.; Zimmermann, J.; Erfle, H.; Paces,
 V.; Ansoerge, W.; Zimmermann, J.; Erfle, H.; Paces,
 Submitted to the EMBL Data Library, August 1995
 Sequencing of 51 kilobases on the right arm of
 from S. cerevisiae reveals 30 open reading frames.
 #accession S60987
 #molecule_type DNA
 #residues 1-1007 #label WIE
 #cross-references EMBL:X90518; NID:g1050808; PID:g1050813
 #authors Benes, V.; Andrade, M.A.; Rechmann, S.; Teodoru, C.; Banreivi,
 A.; Sander, C.; Valencia, A.; Ansoerge, W.; Voss, H.
 Submitted to the EMBL Data Library, December 1995
 Nucleotide sequence and analysis of a 130 kb fragment of
 yeast chromosome XV.
 #accession S61683
 #molecule_type DNA
 #residues 1-1007 #label BEN
 #cross-references EMBL:X94335; NID:g1262139; PID:e217748; PID:g1164971
 #authors Voss, H.; Benes, V.; Rechmann, S.; Teodoru, C.; Schwager, C.;
 Paces, V.; Ansoerge, W.
 Submitted to the Protein Sequence Database, July 1996
 #accession S67012
 #molecule_type DNA
 #residues 1-1007 #label VOS
 #cross-references EMBL:275035; MIPS:YOR127w; NID:g1420329; PID:e252363;
 #experimental_source strain S288C
 #accession S63860
 #authors Wiemann, S.; Rechmann, S.; Benes, V.; Voss, H.; Schwager, C.;
 Vleck, C.; Stegemann, J.; Zimmermann, J.; Erfle, H.; Paces,
 V.; Ansoerge, W.
 Submitted to the EMBL Data Library, August 1995
 Sequencing and analysis of 51 kb on the right arm of
 chromosome XV from Saccharomyces cerevisiae reveals 30 open
 reading frames.
 #cross-references MIMD:97060020
 #accession S63864
 #status Nucleic acid sequence not shown; translation not shown
 #molecule_type DNA
 #residues 1-1007 #label WIV
 #cross-references EMBL:X90518; NID:g1050808; PID:g1050813
 #note the nucleotide sequence was submitted to the EMBL Data
 Library, August 1995
 #accession JC6097
 #authors Chen, G.; Zheng, L.; Chan, C.S.M.
 Submitted to the EMBL Data Library, August 1995
 The LIM domain-containing dbm1 GTPase-activating protein is
 required for normal cellular morphogenesis in Saccharomyces
 cerevisiae.
 #cross-references MIMD:96239492

Thu May 11 06:50:22 2000

terminal half (residue 1-111)

```
#accession      JC6097
#molecule_type DNA
#residues       1-1007 #label CHF
#cross-references GB:U07421; MID:g460972; PID:g460973
```

GENETICS
SGD:RGAL; DBM1; THE1
SGD:S0005653

#gene	##cross-references	MIPS:YOR127W; SGD:YOR127W	binding repeat homology
15R	YOR127W		

```
#note  
CLASSIFICATION  
#superfamily LIM metal-binding repeat homology #label LIM2  
#accession P08679 #database UniProtKB/Swiss-Prot #release 1796
```

```

13-66
FEATURE
#domain LRM meta 112830 #CHECKSUM
#length 1007
SUMMARY

```

Query Match	81.5%	Score 44; DB 2;	Length 1
Best Local Similarity	42.9%	Pred. No. 1.94e+01;	
Mismatches	3; Conservative	4; Mismatches	0; Indels
Gaps	0;		

Macromes

Db	988	MKERNYI	994
		:: :: :	
Ov	363	MNDRSYV	369

RESULT	ENTRY	TITLE
10	SC1174	#type complete
		hypothetical protein YDR379w - yeast (Saccharomyces cerevisiae)

ALTERNATE_NAMES	hypothetical protein D9481.4
ORGANISM	Saccharomyces cerevisiae
DATE	23-Feb-1996
#formal_name	#sequence_revision
01-Mar-1996	01-Mar-1996
#text_change	

DATE 06-FEB-1990

ACCESSIONS S61174

REFERENCE S61159

#authors Ding, H.

submitted to the EMBL Data Library, June 1995

Accession no. S61159, S61174, S61180, S61181, S61182, S61183, S61184, S61185, S61186, S61187, S61188, S61189, S61190, S61191, S61192, S61193, S61194, S61195, S61196, S61197, S61198, S61199, S61200, S61201, S61202, S61203, S61204, S61205, S61206, S61207, S61208, S61209, S61210, S61211, S61212, S61213, S61214, S61215, S61216, S61217, S61218, S61219, S61220, S61221, S61222, S61223, S61224, S61225, S61226, S61227, S61228, S61229, S61230, S61231, S61232, S61233, S61234, S61235, S61236, S61237, S61238, S61239, S61240, S61241, S61242, S61243, S61244, S61245, S61246, S61247, S61248, S61249, S61250, S61251, S61252, S61253, S61254, S61255, S61256, S61257, S61258, S61259, S61260, S61261, S61262, S61263, S61264, S61265, S61266, S61267, S61268, S61269, S61270, S61271, S61272, S61273, S61274, S61275, S61276, S61277, S61278, S61279, S61280, S61281, S61282, S61283, S61284, S61285, S61286, S61287, S61288, S61289, S61290, S61291, S61292, S61293, S61294, S61295, S61296, S61297, S61298, S61299, S61300, S61301, S61302, S61303, S61304, S61305, S61306, S61307, S61308, S61309, S61310, S61311, S61312, S61313, S61314, S61315, S61316, S61317, S61318, S61319, S61320, S61321, S61322, S61323, S61324, S61325, S61326, S61327, S61328, S61329, S61330, S61331, S61332, S61333, S61334, S61335, S61336, S61337, S61338, S61339, S61340, S61341, S61342, S61343, S61344, S61345, S61346, S61347, S61348, S61349, S61350, S61351, S61352, S61353, S61354, S61355, S61356, S61357, S61358, S61359, S61360, S61361, S61362, S61363, S61364, S61365, S61366, S61367, S61368, S61369, S61370, S61371, S61372, S61373, S61374, S61375, S61376, S61377, S61378, S61379, S61380, S61381, S61382, S61383, S61384, S61385, S61386, S61387, S61388, S61389, S61390, S61391, S61392, S61393, S61394, S61395, S61396, S61397, S61398, S61399, S61400, S61401, S61402, S61403, S61404, S61405, S61406, S61407, S61408, S61409, S61410, S61411, S61412, S61413, S61414, S61415, S61416, S61417, S61418, S61419, S61420, S61421, S61422, S61423, S61424, S61425, S61426, S61427, S61428, S61429, S61430, S61431, S61432, S61433, S61434, S61435, S61436, S61437, S61438, S61439, S61440, S61441, S61442, S61443, S61444, S61445, S61446, S61447, S61448, S61449, S61450, S61451, S61452, S61453, S61454, S61455, S61456, S61457, S61458, S61459, S61460, S61461, S61462, S61463, S61464, S61465, S61466, S61467, S61468, S61469, S61470, S61471, S61472, S61473, S61474, S61475, S61476, S61477, S61478, S61479, S61480, S61481, S61482, S61483, S61484, S61485, S61486, S61487, S61488, S61489, S61490, S61491, S61492, S61493, S61494, S61495, S61496, S61497, S61498, S61499, S61500, S61501, S61502, S61503, S61504, S61505, S61506, S61507, S61508, S61509, S61510, S61511, S61512, S61513, S61514, S61515, S61516, S61517, S61518, S61519, S61520, S61521, S61522, S61523, S61524, S61525, S61526, S61527, S61528, S61529, S61530, S61531, S61532, S61533, S61534, S61535, S61536, S61537, S61538, S61539, S61540, S61541, S61542, S61543, S61544, S61545, S61546, S61547, S61548, S61549, S61550, S61551, S61552, S61553, S61554, S61555, S61556, S61557, S61558, S61559, S61560, S61561, S61562, S61563, S61564, S61565, S61566, S61567, S61568, S61569, S61570, S61571, S61572, S61573, S61574, S61575, S61576, S61577, S61578, S61579, S61580, S61581, S61582, S61583, S61584, S61585, S61586, S61587, S61588, S61589, S61590, S61591, S61592, S61593, S61594, S61595, S61596, S61597, S61598, S61599, S61600, S61601, S61602, S61603, S61604, S61605, S61606, S61607, S61608, S61609, S61610, S61611, S61612, S61613, S61614, S61615, S61616, S61617, S61618, S61619, S61620, S61621, S61622, S61623, S61624, S61625, S61626, S61627, S61628, S61629, S61630, S61631, S61632, S61633, S61634, S61635, S61636, S61637, S61638, S61639, S61640, S61641, S61642, S61643, S61644, S61645, S61646, S61647, S61648, S61649, S61650, S61651, S61652, S61653, S61654, S61655, S61656, S61657, S61658, S61659, S61660, S61661, S61662, S61663, S61664, S61665, S61666, S61667, S61668, S61669, S61670, S61671, S61672, S61673, S61674, S61675, S61676, S61677, S61678, S61679, S61680, S61681, S61682, S61683, S61684, S61685, S61686, S61687, S61688, S61689, S61690, S61691, S61692, S61693, S61694, S61695, S61696, S61697, S61698, S61699, S61700, S61701, S61702, S61703, S61704, S61705, S61706, S61707, S61708, S61709, S61710, S61711, S61712, S61713, S61714, S61715, S61716, S61717, S61718, S61719, S61720, S61721, S61722, S61723, S61724, S61725, S61726, S61727, S61728, S61729, S61730, S61731, S61732, S61733, S61734, S61735, S61736, S61737, S61738, S61739, S61740, S61741, S61742, S61743, S61744, S61745, S61746, S61747, S61748, S61749, S61750, S61751,

```
#submission sumit000001
#description The sequence of S. cerevisiae CosmX27
#accession S61174
##molecule_type DNA
##label 1-1009 ##label DIN
##residues 1-1009 ##label DIN
##cross-references EMBL:U2873; NID:g849184; PTD:g849200; MIPS:YDR379W
##cross-references COSMOS:train_S288C (AB972)
```

```
##experimental_source struct
enum RCS
```

GENEIDCS SGD:RGAZ
#gene SGD:S0002787; MIPS:YDR379W
#cross-references repeat homology

```

#map_position 4K
CLASSIFICATION
FEATURE
13-66
#length 1009 #molecular-weight 113290 #checksum 3309
#domain LIM metal-binding repeat homology #label LIM2
#superfamily LIM metal-binding repeat homology #label LIM2

```

SUMMARY	
Query Watch	81.5%; Score 44; DB 2; Length 1009;
Best Local Similarity	42.9%; Pred No. 1.94e+01;
Watches	3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 988 MKERNYI 994
1:1:1:1:
363 MNDRSYV 369

```

RESULT      11          EPBO    #type complete
ENTRY       phosphatidylcholine transfer protein - bovine
TITLE       Bos primigenius taurus #common_name cattle
FORMAL_NAME 31-Aug-1980 #sequence_revision 22-May-1981 #text_change
ORGANISM    31-Aug-1980 #sequence_revision 22-May-1981 #text_change
DATE        31-Aug-1980 #sequence_revision 22-May-1981 #text_change

```

Accessions	Reference	Authors	#Journal	#Title
A9102; A91105; A91260; S68244; A00137		Werner, P.; Akeröy, R.; Westerman, J.; Puyk, W.C.; Smits, J.		The primary structure of the phosphatidylcholine-exchange protein from bovine liver. Isolation and characterization of the staphylococcal protease peptides and the amino-acid

```

sequence of the N-terminal half (residues 1-122)
cross-references MUID:82186640
accession A91092
#molecule_type protein
#residues 1-122 #label MOL
#authors A91105
#journal K.W.A. Biochem. (1981) 114:385-391
#title Eur. J. Biochem. (1981) 114:385-391
#abstract The complete primary structure of the phosphatidylcholine-transfer protein from bovine liver. Isolation and characterization of the cyanogen bromide peptides.
cross-references MUID:81164528
accession A91105
#molecule_type protein
#residues 110-213 #label AKE
#authors A91260 P.; Haagsman, H.P.; van Deenen, L.L.M.; Wirtz, K.W.A.
#journal Moonen, H. (1979) 99:439-445
#title Determination of the hydrophobic binding site of phosphatidylcholine exchange protein with photosensitive phosphatidylcholine.
cross-references MUID:80046712
accession A91260
#molecule_type protein
#residues 145-183 #label MO2
#note residues 171-176 form part of the binding site for phosphatidylcholine
#authors S8244
#journal Geijtenbeek, T.B.H.; Smith, A.J.; Borst, P.; Wirtz, K.W.A.
#title CDNA cloning and tissue-specific expression of the phosphatidylcholine transfer protein gene.
cross-references MUID:96235171
accession S68244
#molecule_type mRNA
#residues 1-166, 'K', 168-213 #label GEI
#cross-references EMBL:D50026; NID:g897692; PIDN:CAA90330.1; PDB:g897693
#note experimental_source liver
#comment The unique property of this protein is to extract phosphatidylcholine from an interface.
#classification superfamily phosphatidylcholine-transfer protein acetylated amino end
#feature
1
17-63,93-207 #modified_site acetylated amino end (Met) #status experimental
#disulfide_bonds #status experimental
#length 213 #molecular_weight 24643 #checksum 4663
79.6% score 43; DB 1; Length 213;
Query Match 57.1%; Pred. No. 3,23e+01;
Best Local Similarity 1; Mismatches 1; Indels 0; Gaps 0;
Matches 4; Conservative
Db 109 MSNRDXY 115
QY 363 MNDSEYV 369
12 E60233 #type complete
RESULT carbamoyl-phosphate synthase, large subunit -
ENTRY Methanobacterium thermoautotrophicum (strain Delta H)
TITLE #formal_name Methanobacterium thermoautotrophicum
ORGANISM 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
DATE 07-Aug-1998
E69233
ACCESSIONS A69000
REFERENCE Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicalaire,

```

Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Petrokovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. (1997) 179:7135-7155
Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.
#accession E69233
#status preliminary; nucleic acid sequence not shown;
translation not shown

##molecule_type DNA
##residues 1-230 #label MTH
##cross-references GB:AE000872; GB:AE000666; NID:g2622098
#experimental_source strain Delta H
#gene MTH997
CLASSIFICATION #superfamily biotin carboxylase homology
9-230
#domain biotin carboxylase homology
#label BCI
#length 230 #molecular-weight 24981 #status atypical

Query Match 79.6%; Score 43; DB 2; Length 2201
Best Local Similarity 71.4%; Pred. No. 3.23e+01;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 60 MADRVV 66
QY 363 MNDRSV 369

RESULT 13
ENTRY
TITLE
ORGANISM
#variety
DATE
ACCESSIONS
REFERENCE
#authors
S75495 #type complete
hypothetical protein slr2119 - Synecocystis sp. (strain PCC 6803)
#formal_name Synecocystis sp.
25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
S75495
S74322
Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Muraki, A.; Nakazaki, N.; Hosouchi, T.; Matsuno, A.; S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. (1996) 3:109-136
Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
#accession S75495
#status preliminary
##molecule_type DNA
##residues 1-278 #label KAN
##cross-references EMBL:D90311; GB:AB001339; NID:g1653083; PID:d1018789;
#note the nucleotide sequence was submitted to the EMBL Data Library, June 1996

SUMMARY
#length 278 #molecular-weight 32463 #checksum 2985
Query Match 79.6%; Score 43; DB 2; Length 278;
Best Local Similarity 42.9%; Pred. No. 3.23e+01;
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 176 MTERGV 182
QY 363 MNDRSV 369

RESULT 14
ENTRY
TITLE

ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:99206606
#accession G72029
#status Preliminary
##molecule_type DNA
##residues 1-441 #label ARN
##cross-references GB:AE001664; GB:AE001363; NID:g4377147; PID:g4377154
#experimental_source strain CWL029
GENETICS
#gene CPn0840
SUMMARY
#length 441 #molecular-weight 50484 #checksum 1815
Query Match 79.6%; Score 43; DB 2; Length 441;
Best Local Similarity 57.1%; Pred. No. 3.23e+01;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 279 MDQDYY 285
QY 363 MNDRSV 369

RESULT 15
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:98196666
#accession F70364
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-465 #label AQF
##cross-references GB:AE000705; NID:g2983310; PIDN:AA06905.1;
#experimental_source strain VF5
GENETICS
#gene lpda
CLASSIFICATION #superfamily dihydroliipoamide dehydrogenase; dihydroliipoamide dehydrogenase homology
FEATURE
6-447
#domain dihydroliipoamide dehydrogenase homology
SUMMARY
#length 465 #molecular-weight 51564 #checksum 9628
Query Match 79.6%; Score 43; DB 2; Length 465;
Best Local Similarity 42.9%; Pred. No. 3.23e+01;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

RESULT 15
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:98196666
#accession F70364
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-465 #label AQF
##cross-references GB:AE000705; NID:g2983310; PIDN:AA06905.1;
#experimental_source strain VF5
GENETICS
#gene lpda
CLASSIFICATION #superfamily dihydroliipoamide dehydrogenase; dihydroliipoamide dehydrogenase homology
FEATURE
6-447
#domain dihydroliipoamide dehydrogenase homology
SUMMARY
#length 465 #molecular-weight 51564 #checksum 9628
Query Match 79.6%; Score 43; DB 2; Length 465;
Best Local Similarity 42.9%; Pred. No. 3.23e+01;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

RESULT 15
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:98196666
#accession F70364
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-465 #label AQF
##cross-references GB:AE000705; NID:g2983310; PIDN:AA06905.1;
#experimental_source strain VF5
GENETICS
#gene lpda
CLASSIFICATION #superfamily dihydroliipoamide dehydrogenase; dihydroliipoamide dehydrogenase homology
FEATURE
6-447
#domain dihydroliipoamide dehydrogenase homology
SUMMARY
#length 465 #molecular-weight 51564 #checksum 9628
Query Match 79.6%; Score 43; DB 2; Length 465;
Best Local Similarity 42.9%; Pred. No. 3.23e+01;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

RESULT 15
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:98196666
#accession F70364
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-465 #label AQF
##cross-references GB:AE000705; NID:g2983310; PIDN:AA06905.1;
#experimental_source strain VF5
GENETICS
#gene lpda
CLASSIFICATION #superfamily dihydroliipoamide dehydrogenase; dihydroliipoamide dehydrogenase homology
FEATURE
6-447
#domain dihydroliipoamide dehydrogenase homology
SUMMARY
#length 465 #molecular-weight 51564 #checksum 9628
Query Match 79.6%; Score 43; DB 2; Length 465;
Best Local Similarity 42.9%; Pred. No. 3.23e+01;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

RESULT 15
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:98196666
#accession F70364
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-465 #label AQF
##cross-references GB:AE000705; NID:g2983310; PIDN:AA06905.1;
#experimental_source strain VF5
GENETICS
#gene lpda
CLASSIFICATION #superfamily dihydroliipoamide dehydrogenase; dihydroliipoamide dehydrogenase homology
FEATURE
6-447
#domain dihydroliipoamide dehydrogenase homology
SUMMARY
#length 465 #molecular-weight 51564 #checksum 9628
Query Match 79.6%; Score 43; DB 2; Length 465;
Best Local Similarity 42.9%; Pred. No. 3.23e+01;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

US-09-376-430-2-25.rpr

Thu May 11 06:50:22 2000

Db 283 MDERGV 289
1:1:1
QY 363 MNDRSYV 369

Search completed: Wed May 10 14:30:09 2000
Job time : 8 secs.

DR HSSP; P14336; ISVB.
 DR PFAM; PF00869; Flavi-glycoprot; 1.
 DR PFAM; PF00948; Flavi_NSI; 1.
 DR PFAM; PF00949; Flavi_helicase; 1.
 DR PFAM; PF00972; Flavi_NSS; 1.
 DR PFAM; PF01002; Flavi_NS2B; 1.
 DR PFAM; PF01003; Flavi_Capsid; 1.
 DR PFAM; PF01004; Flavi_M; 1.
 DR PFAM; PF01005; Flavi_NS2A; 1.
 DR PFAM; PF01349; Flavi_NS4B; 1.
 DR PFAM; PF01350; Flavi_NS4A; 1.
 DR PFAM; PF01570; Flavi_Prop; 1.
 KW Polyprotein; Glycoprotein; Transfrase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1 112
 FT CHAIN 113 205
 FT CHAIN 206 280
 FT CHAIN 281 776
 FT CHAIN 777 1128
 FT CHAIN 1129 1358
 FT CHAIN 1359 1489
 FT CHAIN 1490 2110
 FT CHAIN 2111 2259
 FT CHAIN 2260 2511
 FT CHAIN 2512 3414
 FT NP_BIND 1688 1695
 FT SITE 1779 1782
 FT TRANSMEM 103 119
 FT TRANSMEM 262 278
 FT TRANSMEM 728 744
 FT TRANSMEM 758 774
 FT DISULFID 283 310
 FT DISULFID 340 366
 FT DISULFID 372 401
 FT DISULFID 465 570
 FT CARBOHYD 587 618
 FT CARBOHYD 144 144
 FT CARBOHYD 434 434
 SQ SEQUENCE 3414 AA; 378017 MW; 59CB7E95DD70D82E CRC64;
 Query Match 85.2%; Score 46; DB 1; Length 3414;
 Best Local Similarity 57.1%; Pred. No. 1.96e+00;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Db 2576 MEERGTV 2582
 QY 363 MNDRSYV 369
 RESULT 2
 ID VG39_HSV11 STANDARD; PRT; 1123 AA.
 AC Q00143;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DE HYPOTHETICAL GENE 39 PROTEIN.
 GN 39.
 OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae.
 OC unclassified Herpesviridae.
 RN [1]
 RC SEQUENCE FROM N.A.
 RA STRAIN-AUBURN 1;
 RA MEDLINE; 92087490.
 RA Davison A.J.;
 RT "Channel catfish virus: a new type of herpesvirus";
 RL Virology 186:9-14(1992).
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL; M75136; AAA88142.1;
 DR PIR; D36790; D36790.
 KW Hypothetical protein.
 SQ SEQUENCE 1123 AA; 123008 MW; 47953BEB727EBD00 CRC64;
 Query Match 83.3%; Score 45; DB 1; Length 1123;
 Best Local Similarity 83.3%; Pred. No. 3.52e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 534 MNDRAV 539
 QY 363 MNDRSYV 368
 RESULT 3
 ID RGAL_YEAST STANDARD; PRT; 1007 AA.
 AC P39083; P39934;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RHO-TYPE GTPASE ACTIVATING PROTEIN RGAL/DBM1.
 GN RGAL OR DBM1 OR THE1 OR YOR127W OR O3290 OR YOR3290W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RC SEQUENCE FROM N.A., AND MUTAGENESIS.
 RA STRAIN-S288C;
 RA MEDLINE; 96239492.
 RA Chen G., Zheng L., Chan C.S.M.;
 RT "The LIM domain-containing Dbm1 GTPase-activating protein is required
 RT for normal cellular morphogenesis in Saccharomyces cerevisiae";
 RL Mol. Cell. Biol. 16:1376-1390(1996).
 RN [2]
 RC SEQUENCE FROM N.A.
 RA STRAIN-S288C / FY1679;
 RA MEDLINE; 97060020.
 RA Wiemann S., Rechmann S., Voss H., Schwager C., Vlcek C.,
 RT "Sequencing and analysis of 51 kb on the right arm of chromosome XV
 RT from Saccharomyces cerevisiae reveals 30 open reading frames";
 RL Yeast 12:281-288(1996).
 RN [3]
 RC SEQUENCE FROM N.A.
 RA MEDLINE; 97344368.
 RA Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C.,
 RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV";
 RL Yeast 13:655-672(1997).
 RN [4]
 RC SEQUENCE FROM N.A.
 RA STRAIN-S288C;
 RA MEDLINE; 96101594.
 RA Stevenson B.J., Ferguson B., de Virgilio C., Bi E., Pringle J.R.,
 RT "Mutation of RGAL, which encodes a putative GTPase-activating protein
 RT for the polarity-establishment protein Cdc42p, activates the
 RL pheromone-response pathway in the yeast Saccharomyces cerevisiae";
 RL Genes Dev. 9:2949-2963(1995).
 RN [5]
 RC SEQUENCE OF 570-639 FROM N.A.
 RA STRAIN-SNY243;
 RA MEDLINE; 93087574.
 RA Ramer S.W., Ellledge S.J., Davis R.W.;
 RT "Dominant genetics using a yeast genomic library under the control of


```

DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-FEB-2000 (Rel. 39, Last annotation update)
DE PUTATIVE ABC TRANSPORTER ATP-BINDING PROTEIN EGO (EG010A).
GN EGO.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
RA Ikemoto K., Inada T., Isono K., Itoh T., Kanai K., Kasai H.,
RA Masuda S., Miki T., Mizobuchi K., Morikawa M., Nakamura K.,
RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sampaio G., Seki Y.,
RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Kwon H.B., Lee S.H., Choe M.H.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM EGO.YDEVZ. PROBABLY RESPONSIBLE FOR THE ENERGY COUPLING TO
CC THE TRANSPORT SYSTEM. THIS PROTEIN IS ESSENTIAL FOR AEROBIC
CC GROWTH.
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). MGLA/RBSA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000249; AAC74586.1;
DR EMBL; D90794; CAB21119.1;
DR EMBL; AF089855; AAC61747.1;
DR EGOGENE; EG13806; EGO.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR PFAM; PF00005; ABC_TRAN; 2.
KW Transport; ATP-binding.
FT NP_BIND 44 51
SQ SEQUENCE 511 AA; 55821 MW; CC399008BFDD55FA CRC64;

Query Match 79.68; Score 43; DB 1; Length 511;
Best Local Similarity 71.48; Pred. No. 1.10e+01;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 469 MADRVV 475
QY 363 MNDRSYV 369
| | | | |

RESULT 6
ID HEMA_MEASH STANDARD; PRT; 617 AA.
AC P06830;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DE 01-APR-1993 (Rel. 25, Last annotation update)
DE HEMAGGLUTININ-NEURAMINIDASE (EC 3.2.1.18).

Query Match 79.68; Score 43; DB 1; Length 617;
Best Local Similarity 71.48; Pred. No. 1.10e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 31 MIDRPYV 37
QY 363 MNDRSYV 369
| | | | |

RESULT 7
ID HEMA_MEASE STANDARD; PRT; 617 AA.
AC P08302;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE HEMAGGLUTININ-NEURAMINIDASE (EC 3.2.1.18).
GN HN.
OS Measles virus (strain Edmonston) (Subacute sclerosing panencephalitis
OC virus); ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Morbillivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86181590.
RA Alkhatib G., Briedis D.J.
RT "The predicted primary structure of the measles virus hemagglutinin."
RL Virology 150:475-490(1986).
RN [2]

```


SEQUENCE FROM N.A.

EMBL: 90085790.
MEDLINE: 90085790.
RA Ratzan R., Schmid A., Spielhofer P., Kaelin K., Bacsko K.,
Meulen V., Pardowitz J., Flanagan S., Rima B.K., Udem S.A.;
caused lethal human brain diseases.";
RT "Mutated and hypermutated genes of persistent measles viruses
Virology 173:415-425(1989).
RL FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS
TO CELL RECEPTORS AND FOR INITIATING INFECTION. NEURAMINIDASE
CC ACTIVITY HELPS THE EFFICIENT SPREAD OF THE VIRUS BY DISSOCIATING
THE MATURE VIRIONS FROM THE NEURAMINIC ACID CONTAINING
CC GLYCOPROTEINS.
CC SUBCELLULAR LOCATION: EXTERNAL, ANCHORED TO THE ENVELOPE BY ITS
CC N-TERMINAL HYDROPHOBIC SEQUENCE.
CC -!- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES HEMAGGLUTININ-
NEURAMINIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entitles requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL: S58435; AAB26146.1; .
PIR: F48556; F48556.
PFAM: PF00423; HN; 1.
KW Hydrolase; Hemagglutinin; Envelope protein; Glycoprotein;
DOMAIN 35 58 CYTOPLASMIC (POTENTIAL).
TRANSMEM 1 34 MEMBRANE ANCHOR (POTENTIAL).
FT TRANSMEM 35 617 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 168 187 POTENTIAL.
FT CARBOHYD 187 187 POTENTIAL.
FT CARBOHYD 200 200 POTENTIAL.
FT CARBOHYD 215 215 POTENTIAL.
FT CARBOHYD 238 238 POTENTIAL.
SQ SEQUENCE 617 AA; 59249 MW; 0EFA05AEDA43D9C6 CRC64;

Query Match Score 43; DB 1; Length 617;
Best Local Similarity 71.4%; Pred. No. 1.10e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 31 MIDRPVV 37
363 MNDRSYV 369

RESULT 8 STANDARD; PRT; 617 AA.
ID HEMA_MEASA
AC P35971; 1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HEMAGGLUTININ-NEURAMINIDASE (EC 3.2.1.18).
GN Measles virus (strain AIK-C) (Subacute sclerosing panencephalitis
virus); ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Morbillivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93227570.
RA Mori T., Sasaki K., Hashimoto H., Makino S.;
Molecular cloning and complete nucleotide sequence of genomic RNA of
the AIK-C strain of attenuated measles virus.";
RT Virus Genes 7:67-81(1993).
RL FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS
TO CELL RECEPTORS AND FOR INITIATING INFECTION. NEURAMINIDASE
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS
CC ACTIVITY HELPS THE EFFICIENT SPREAD OF THE VIRUS BY DISSOCIATING
CC THE MATURE VIRIONS FROM THE NEURAMINIC ACID CONTAINING

SEQUENCE FROM N.A.

EMBL: M14877; AAA46424.1; .
EMBL: K01711; AAA75500.1; .
PIR: A27006; HMNZED.
PFAM: PF00423; HN; 1.
KW Hydrolase; Hemagglutinin; Envelope protein; Glycoprotein;
Transmembrane. 1 34 CYTOPLASMIC (POTENTIAL).
DOMAIN 35 58 MEMBRANE ANCHOR (POTENTIAL).
TRANSMEM 59 617 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 168 187 POTENTIAL.
FT DOMAIN 187 187 POTENTIAL.
FT CARBOHYD 200 200 POTENTIAL.
FT CARBOHYD 215 215 POTENTIAL.
FT CARBOHYD 238 238 POTENTIAL.
SQ SEQUENCE 617 AA; 59249 MW; 0EFA05AEDA43D9C6 CRC64;

Query Match Score 43; DB 1; Length 617;
Best Local Similarity 71.4%; Pred. No. 1.10e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 31 MIDRPVV 37
363 MNDRSYV 369

RESULT 8 STANDARD; PRT; 617 AA.
ID HEMA_MEASA
AC P35971; 1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HEMAGGLUTININ-NEURAMINIDASE (EC 3.2.1.18).
GN Measles virus (strain AIK-C) (Subacute sclerosing panencephalitis
virus); ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Morbillivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93227570.
RA Mori T., Sasaki K., Hashimoto H., Makino S.;
Molecular cloning and complete nucleotide sequence of genomic RNA of
the AIK-C strain of attenuated measles virus.";
RT Virus Genes 7:67-81(1993).
RL FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS
TO CELL RECEPTORS AND FOR INITIATING INFECTION. NEURAMINIDASE
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS
CC ACTIVITY HELPS THE EFFICIENT SPREAD OF THE VIRUS BY DISSOCIATING
CC THE MATURE VIRIONS FROM THE NEURAMINIC ACID CONTAINING

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D10549; BAA01406.1; --
PIR; J00273; J00273.
PFAM; PF00423; HN; 1.
Hydrolase; Hemagglutinin; Envelope protein; Glycoprotein;
Transmembrane.
DOMAIN 1 34
FT TRANSMEM 35 58
FT DOMAIN 59 620
FT CARBOHYD 168 168
FT CARBOHYD 187 187
FT CARBOHYD 200 200
FT CARBOHYD 215 215
FT CARBOHYD 238 238
SQ SEQUENCE 620 AA; 69601 MW; 30641B0DA2B48A68 CRC64;

Query Match
Best Local Similarity 79.6%; Score 43; DB 1; Length 620;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 31 MIDRPV 37
QY 363 MNDRSYV 369

RESULT 10
ID PLO2_HUMAN STANDARD; PRT; 737 AA.
AC O00469;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE PROCOLLAGEN-LYSINE,2-OXOGLUTARATE 5-DIOXYGENASE 2
GN PLOD2.1 (Lysyl Hydroxylase 2) (LH2).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97207229.
RA Valtavaara M., Paipponen H., Pirttila A.M., Hiltunen K., Helander H.,
RT "Cloning and characterization of a novel human lysyl hydroxylase
isoform highly expressed in pancreas and muscle."
RL J. Biol. Chem. 272:6831-6834 (1997).
CC -!- FUNCTION: FORMS HYDROXYLYSINE RESIDUES IN -XAA-LYS-GLY- SEQUENCES
FOR CARBOHYDRATE UNITS AND ARE ESSENTIAL FOR THE STABILITY OF THE
INTERMOLECULAR COLLAGEN CROSSLINKS.
CC -!- CATALYTIC ACTIVITY: PROCOLLAGEN L-LYSINE + 2-OXOGLUTARATE + O(2) -
PROCOLLAGEN 5-HYDROXY-L-LYSINE + SUCCINATE + CO(2).
CC -!- COFACTOR: REQUIRES IRON AND ASCORBATE (BY SIMILARITY).
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: MEMBRANE BOUND IN CISTERNAE OF ROUGH
ENDOPLASMIC RETICULUM.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PANCREAS AND MUSCLE.
CC -!- SIMILARITY: BELONGS TO THE LYSYL HYDROXYLASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U84573; AAB58363.1; --
MIM; 601865; --
DR PROSITE; PS01325; LYS_HYDROXYLASE; 1.
KW Oxidoreductase; Dioxxygenase; Signal; Iron; Vitamin C; Glycoprotein;
Endoplasmic reticulum; Membrane.

FT SIGNAL 1 25
FT CHAIN 26 737
FT METAL 666 666
FT METAL 668 668
FT ACT SITE 718 718
FT CARBOHYD 728 728
FT CARBOHYD 63 63
FT CARBOHYD 209 209
FT CARBOHYD 297 297
FT CARBOHYD 365 365
FT CARBOHYD 522 522
FT CARBOHYD 598 598
FT CARBOHYD 696 696
FT CARBOHYD 725 725
SQ SEQUENCE 737 AA; 84663 MW; 09AB83013AFBB002 CRC64;

Query Match
Best Local Similarity 79.6%; Score 43; DB 1; Length 737;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 476 MNERNY 481
QY 363 MNDRSYV 368

RESULT 11
ID SYL_RICPR STANDARD; PRT; 828 AA.
AC Q92DB1;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DE PROCOLLAGEN-LYSINE,2-OXOGLUTARATE 5-DIOXYGENASE 2
GN LEUCYL-TRNA SYNTHETASE (EC 6.1.1.4) (LEUCINE--TRNA LIGASE) (LEURS).
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsia.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-MADRID E;
RX MEDLINE; 99039499.
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RT "The genome sequence of Rickettsia prowazekii and the origin of
mitochondria."
RL Nature 396:133-140 (1998).
CC -!- CATALYTIC ACTIVITY: ATP + L-LEUCINE + TRNA(LEU) - AMP +
PYROPHOSPHATE + L-LEUCYL-TRNA(LEU).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; A7235271; CAAL4878.1;
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT SIMILAR 35 45
FT SIMILAR 595 599
FT BINDING 598 598
FT BINDING 598 598
SQ SEQUENCE 828 AA; 96471 MW; 43BACB6398191986 CRC64;

Query Match
Best Local Similarity 79.6%; Score 43; DB 1; Length 828;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 533 MNERNYV 539
QY 363 MNDRSYV 368

US-09-376-430-2-25.rsp

Thu May 11 06:50:22 2000

Query Match 77.8%; Score 42; DB 1; Length 124;
Best Local Similarity 83.3%; Pred. No. 1.92e+01;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 107 MODRSY 112
QY 363 MNDRSY 368

Search completed: Wed May 10 14:25:29 2000
Job time : 97 secs.

VARIANT 1227 1227 P -> L (IN TSC).
FT FTID=VAR_005657.
VARIANT 1240 1240 R -> W (IN TSC).
FT FTID=VAR_005658.
VARIANT 1295 1295 D -> V (IN TSC).
FT FTID=VAR_005659.
VARIANT 1315 1315 P -> S (IN TSC).
FT FTID=VAR_008028.
VARIANT 1329 1329 R -> H (IN TSC).
FT FTID=VAR_008029.
VARIANT 1509 1509 MISSING (IN TSC).
FT FTID=VAR_005660.
VARIANT 1549 1549 Y -> C (IN TSC).
FT FTID=VAR_005661.
VARIANT 1614 1614 MISSING (IN TSC).
FT FTID=VAR_005662.
VARIANT 1643 1643 N -> I (IN TSC).
FT FTID=VAR_005663.
VARIANT 1650 1650 Y -> C (IN TSC).
FT FTID=VAR_005664.
VARIANT 1690 1690 D -> Y (IN TSC).
FT FTID=VAR_005665.
VARIANT 1709 1709 P -> L (IN TSC).
FT FTID=VAR_008030.
VARIANT 1712 1712 A -> E (IN TSC).
FT FTID=VAR_005666.
FT

...Note: remainder of annotations omitted.

Query Match 79.6%; Score 43; DB 1; Length 1807;
Best Local Similarity 57.1%; Pred. No. 1.10e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 280 MEDRAYM 286
QY 363 MNDRSYV 369

RESULT 15
ID YHCC_BACSU STANDARD; PRT; 124 AA.
AC P54587;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DT HYPOTHETICAL 14.0 KD PROTEIN IN GLPD-CSPB INTERGENIC REGION.
DE YHCC.
GN Bacillus subtilis.
OS Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-168;
RC MEDLINE; 97124185.
RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RT "A 22 kb DNA sequence in the cspB-glpPKD region at 75 degrees on the
Bacillus subtilis chromosome."
RL Microbiology 142:3021-3025(1996).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X96983; CAA65686.1;
DR EMBL; Z99108; CAB12731.1;
DR SUBTILIST; BG11581; yhcc.
KW HYPOTHETICAL protein; Transmembrane.
KW TRANSHEM 2 22 POTENTIAL.
SQ SEQUENCE 124 AA; 13958 MW; F9050B1152A38F3C CRC64;

RA LI M.J., LIU C.R., BRYANT Z., SUBRAHMANYAN L., RUOHOLA-BAKER H.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF093637; AAC72365.1; -
 DR HSSP: P28867; 1PTR.
 SQ SEQUENCE 1750 AA; 198941 MW; 385F5B68 CRC32;

Query Match 87.0%; Score 47; DB 5; Length 1750;
 Best Local Similarity 71.4%; Pred. No. 3.35e+00;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 913 MRERSV 919
 | :|||
 QY 363 MDRSV 369

RESULT 3
 ID O83178 PRELIMINARY; PRT; 220 AA.
 AC O83178;
 DT 01-NOV-1998 (TREMREL. 08, Created)
 DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
 DT 01-MAY-1999 (TREMREL. 10, Last annotation update)
 DE THIAMINE ABC TRANSPORTER, ATP-BINDING PROTEIN, PUTATIVE.
 GN TP0142.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98322770.
 RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
 RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
 RA SODERGEN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
 RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
 RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
 RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
 RA VENTER J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 RT spirochete";
 RL Science 281:375-388(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98322770.
 RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
 RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
 RA SODERGEN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
 RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
 RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
 RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
 RA VENTER J.C.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE001199; AAC65130.1; -
 DR TIGR: TP0142;
 DR PFAM: PF00005; ABC_tran; 1.
 KW ATP-binding.
 SQ SEQUENCE 220 AA; 24066 MW; F13CC6A1 CRC32;

Query Match 85.2%; Score 46; DB 2; Length 220;
 Best Local Similarity 57.1%; Pred. No. 5.89e+00;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1 MDRPV 7
 | :|||
 QY 363 MDRSV 369

RESULT 4
 ID P73650 PRELIMINARY; PRT; 240 AA.
 AC P73650;
 DT 01-FEB-1997 (TREMREL. 02, Created)
 DT 01-FEB-1997 (TREMREL. 02, Last sequence update)
 DT 01-NOV-1999 (TREMREL. 12, Last annotation update)
 DE HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT ATP-BINDING PROTEIN
 GN BRAG.
 OS Synechocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RA TABATA S.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RX MEDLINE: 97061201.
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
 RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
 RA TABATA S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 DR EMBL: D90908; BAAL7695.1; -
 DR HSSP: P13569; INBD.
 DR PFAM: PF00005; ABC_tran; 1.
 KW ATP-binding.
 SQ SEQUENCE 240 AA; 25705 MW; 3AF7879D CRC32;

Query Match 85.2%; Score 46; DB 2; Length 240;
 Best Local Similarity 71.4%; Pred. No. 5.89e+00;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 202 MADRGV 208
 | :|||
 QY 363 MDRSV 369

RESULT 5
 ID O32180 PRELIMINARY; PRT; 110 AA.
 AC O32180;
 DT 01-JAN-1998 (TREMREL. 05, Created)
 DT 01-JAN-1998 (TREMREL. 05, Last sequence update)
 DT 01-NOV-1998 (TREMREL. 08, Last annotation update)
 DE YUSN PROTEIN.
 GN YUSN.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE: 98044033.
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
 RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
 RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
 RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMINGS N.J., DANIEL R.A.,
 RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
 RA ETIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
 RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
 RA GHIM S.Y., GLASER P., GORFEAU A., GOLIGHTLY E.J., GRANDI G.,
 RA GUISEPPI G., GUY B.J., HAGA K., HALECH J., HARWOOD C.R., HENAUT A.,
 RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
 RA JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
 RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUNANO M.,
 RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
 RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTIL D., NAKAI S., NOBACK M.,
 RA NOONE D., O'REILLY M., OGAWA K., OGIMARA A., OUDEGA B., PARK S.H.,
 RA PARRO V., POHL T.M., PORTELLE D., PORWOLLIK S., PRESCOTT A.M.,
 RA PRESCAN E., POUTIC P., PURNELLE B., RAPORT G., REY M., REYNOLDS S.,
 RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE F.,
 RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
 RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDI B.,
 RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
 RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,

US-09-376-430-2-25.rspt

Thu May 11 06:50:22 2000

```

RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA VIARI A., WAMBUTT R., WEDLER H., WEITZENEGGER T.,
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN-168;
RC KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RL EMBL: Z99120; CAB15275.1; -
DR EMBL: 110 AA; 13138 MW; 4D7F15B6 CRC32;
SQ SEQUENCE 110 AA; 13138 MW; 4D7F15B6 CRC32;

Query Match 83.3%; Score 45; DB 2; Length 110;
Best Local Similarity 71.4%; Pred. No. 1.03e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 21 MNDREFV 27
| | | | |
QY 363 MNDRSYV 369

RESULT 6 PRELIMINARY; PRT: 251 AA.
ID O61607
AC O61607;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE F12F6.6 (FRAGMENT).
OS Drosophila heteroneura (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RN SEQUENCE FROM N.A.
RP DAVIS T., KURIHARA J., YAMAMOTO D.;
RA Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF053310; AAC08573.1; -
DR FLYBASE; Fbgn0024446; Dhet\F12F6.6.
FT NON_TER 1
FT NON_SEQUENCE 251 AA; 28591 MW; 9C077574 CRC32;
SQ SEQUENCE 251 AA; 28591 MW; 9C077574 CRC32;

Query Match 83.3%; Score 45; DB 5; Length 251;
Best Local Similarity 71.4%; Pred. No. 1.03e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 83 LDRSYV 89
| | | | |
QY 363 MNDRSYV 369

RESULT 7 PRELIMINARY; PRT: 1687 AA.
ID Q9WUF6
AC Q9WUF6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE TUBEROUS SCLEROSIS 2 PROTEIN (FRAGMENT).
OS Mus musculus domesticus (western European house mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RN SEQUENCE FROM N.A.
RP KLEYMENOVA E.V., DECLUE J., WALKER C.L.;
RA "Genetic variants of the tuberous sclerosis tumor suppressor gene in
RT mouse t haplotypes."
RL Genet. Res. 0:0-0(1999).
DR EMBL: AF132986; AAD27867.1; -
FT NON_TER 1

TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
VIARI A., WAMBUTT R., WEDLER H., WEITZENEGGER T.,
WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
"The complete genome sequence of the gram-positive bacterium Bacillus
subtilis."
Nature 390:249-256(1997).
[2]
SEQUENCE FROM N.A.
STRAIN-168;
KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
EMBL: Z99120; CAB15275.1; -
EMBL: 110 AA; 13138 MW; 4D7F15B6 CRC32;

Query Match 83.3%; Score 45; DB 11; Length 1687;
Best Local Similarity 71.4%; Pred. No. 1.03e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 162 MEDRSYM 168
| | | | |
QY 363 MNDRSYV 369

RESULT 8 PRELIMINARY; PRT: 1814 AA.
ID Q61037
AC Q61037; Q61007; Q61008; P97723; P97724; P97725; P97727;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE TUBERIN (TUBEROUS SCLEROSIS 2 PROTEIN).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RN SEQUENCE FROM N.A. (VARIANTS A-G).
RP TISSUE=HEART;
RC MEDLINE; 96258425.
RA KIM K.K., PAJAK L., WANG H., FIELD L.J.;
RT "Cloning, developmental expression, and evidence for alternative
RT splicing of the murine tuberous sclerosis (TSC2) gene product.";
RL Cell. Mol. Biol. Res. 41:515-526(1995).
RN [2]
RN SEQUENCE FROM N.A. (VARIANTS H AND I).
RX MEDLINE; 96430093.
RA OLSSON P.G., SCHOFIELD J.N., EDWARDS Y.H., FRISCHAUF A.M.;
RT "Expression and differential splicing of the mouse TSC2 homolog.";
RL Mamm. Genome 7:212-215(1996).
RN [3]
RN SEQUENCE OF 1-199 FROM N.A.
RP STRAIN-BALB/C; TISSUE=BLOOD;
RC SARKER A.H., IKEDA S., NAKANO H., TERATO H., IDE H., IMAI K.,
RA AKIYAMA K., TSUTSUI K., BO Z., KOBO K., YAMAMOTO K., YASUI A.,
RA YOSHIDA M., SEKI S.;
RT "Cloning and characterization of a mouse homologue (mNth11) of
RT Escherichia coli endonuclease III.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: SPECIFICALLY STIMULATES THE INTRINSIC GTPASE ACTIVITY OF
CC THE RAS-RELATED PROTEIN RAPIA. SUGGESTING A POSSIBLE MECHANISM FOR
CC ITS ROLE IN REGULATING CELLULAR GROWTH. MUTATIONS IN TUBERIN LEADS
CC TO CONSTITUTIVE ACTIVATION OF RAPIA IN TUMORS.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 9 ISOFORMS OF THE PROTEIN (VARIANTS
CC A-I) CAN BE DERIVED FROM ALTERNATIVE SPLICING OF THE TSC2 GENE.
CC THE SEQUENCE SHOWN HERE IS THAT OF VARIANT G.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC -!- SIMILARITY: TO RAPICAP.
DR EMBL: U37775; AAA86902.1; -
DR EMBL: U39818; AAB18754.1; -
DR EMBL: U37775; AAA86901.1; -
DR EMBL: AB009371; BAA28845.1; -
DR MGD; MGI:102548; Tsc2.
KW Anti-oncogene; Alternative splicing; GTPase activation.
FT VARSPLIC 79 115 MISSING (IN VARIANT A).
FT VARSPLIC 534 572 MISSING (IN VARIANT B).
FT VARSPLIC 946 988 MISSING (IN VARIANT H).
FT VARSPLIC 947 990 MISSING (IN VARIANT C).
FT VARSPLIC 1245 1258 GHAPQVIVSATGC -> RDTALYKSLSPVPAAG (IN
FT VARIANTS H AND I).
FT VARSPLIC 1271 1293 MISSING (IN VARIANT D).
FT VARSPLIC 1281 1303 MISSING (IN VARIANTS H AND I).
FT VARSPLIC 1692 1728 DGPACKEWNRQPGIIVWALPVMELTIVILLCHLQ ->
DMEGLVDYTSVAKIVSDNLSFVAROMALHAN (IN
FT VARIANTS E).
FT VARSPLIC 1693 1728 GPACKCEWNRQPGIIVWALPVMELTIVILLCHLQ -> M

```

FT EGVLTSVAKIVSDRLMSFVARQWALHAN (IN
 FT VARIANTS H AND I).
 FT IREVVHNSPLPL -> VGRTRFMSMSG (IN
 FT VARIANT F).
 FT VARSPLIC 1762 1775
 FT VARSPLIC 1776 1814
 SQ MISSING (IN VARIANT F).
 SQ SEQUENCE 1814 AA; 202069 MW; 89CCCF38 CRC32;

Query Match 83.38; Score 45; DB 11; Length 1814;
 Best Local Similarity 71.4%; Pred. No. 1.03e+01;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 280 MDRSYN 286
 QY 363 MDRSYV 369

RESULT 9
 ID Q69356 PRELIMINARY; PRT; 180 AA.
 AC Q69356;
 DT 01-NOV-1996 (TREMREL. 01, Created)
 DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
 DT 01-JAN-1999 (TREMREL. 09, Last annotation update)
 DE HOMOLOGUE OF HSV-1 GG (FRAGMENT).
 OS Feline herpesvirus (Feline herpesvirus 1).
 OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicelloviruses.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-G2620;
 RX MEDLINE: 95266277.
 RA WILLESE M.J., STRIJVDEN I.G., VAN SCHOONEVELD S.H.,
 RA VAN DEN BERG M.C., SONDERMEIJER P.J.;
 RT "transcriptional analysis of the short segment of the feline
 RT herpesvirus type 1 genome and insertional mutagenesis of a unique
 RT reading frame";
 RL Virology 208:704-711(1995).
 DR EMBL: D42113; BAA07690.1; -;
 FT NON TER 1
 SQ SEQUENCE 180 AA; 20223 MW; 7F475BB3 CRC32;

Query Match 81.5%; Score 44; DB 14; Length 180;
 Best Local Similarity 71.4%; Pred. No. 1.77e+01;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 154 MDRSYV 160
 QY 363 MDRSYV 369

RESULT 10
 ID Q31853 PRELIMINARY; PRT; 287 AA.
 AC Q31853;
 DT 01-JAN-1998 (TREMREL. 05, Created)
 DT 01-JAN-1998 (TREMREL. 05, Last sequence update)
 DT 01-NOV-1998 (TREMREL. 08, Last annotation update)
 DE YOJK PROTEIN.
 GN YOJK.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE: 98044033.
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
 RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
 RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
 RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMINGS N.J., DANIEL R.A.,
 RA DENIZOT F., DEVINE K.M., DUSTERHOFF A., EHRLICH S.D., EMERSON P.T.,
 RA ENTIAN K.D., ERINGTON J., FABRET C., FERRARI E., FOULGER D.,
 RA FRITZ C., FUJITA M., FUJITA Y., FUNA S., GALIZZI A., GALLERON N.,
 RA GHM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,

RA GUISEPPI G., GUY B.J., HAGA K., HATECH J., HARWOOD C.R., HENAUT A.,
 RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
 RA JORIS B., KARAMATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,
 RA KOBAYASHI Y., KOETER P., KONINGSTEIN G., KROGH S., KUNANO M.,
 RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
 RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
 RA NOONE D., O'REILLY M., OGAWA K., OGAWA A., ODEGA B., PARK S.H.,
 RA PARRO V., POHL T.M., PORTELELLI D., FORMOLLI S., PRESCOTT A.M.,
 RA PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
 RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADATE Y.,
 RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
 RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDI B.,
 RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
 RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
 RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
 RA VIARI A., WAMBUUT R., WEDLER E., WEDLER H., WEITZENEGER T.,
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis";
 RL Nature 390:249-256(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA PARK S.-H., SHIN B.-S., CHOI S.-K., GHIM S.-Y.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z99114; CAB3834.1; -;
 DR EMBL: AF026147; AAC17859.1; -;
 DR PFAM: PF00201; UDPGT; 1.
 SQ SEQUENCE 287 AA; 31990 MW; 38117D25 CRC32;

Query Match 81.5%; Score 44; DB 2; Length 287;
 Best Local Similarity 83.3%; Pred. No. 1.77e+01;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 245 MDRSYV 250
 QY 363 MDRSYV 368

RESULT 11
 ID Q9XBH5 PRELIMINARY; PRT; 393 AA.
 AC Q9XBH5;
 DT 01-NOV-1999 (TREMREL. 12, Created)
 DT 01-NOV-1999 (TREMREL. 12, Last sequence update)
 DT 01-NOV-1999 (TREMREL. 12, Last annotation update)
 DE CHROMATE TRANSPORT PROTEIN.
 GN CHRA.
 OS Bacillus cereus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 10987;
 RX MEDLINE: 99231848.
 RA OKSTAD O.A., HEGNA I., LINDBAECK T., RISHOVD A.L., KOLSTO A.B.;
 RT "Genome organization is not conserved between Bacillus cereus and
 RT Bacillus subtilis";
 RL Microbiology 145:621-631(1999).
 DR EMBL: AJ000394; CAB40623.1; -;
 SQ SEQUENCE 393 AA; 42860 MW; F3DBE8AD CRC32;

Query Match 81.5%; Score 44; DB 2; Length 393;
 Best Local Similarity 66.7%; Pred. No. 1.77e+01;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 47 MDRSYV 52

US-09-376-430-2-25.ispt

Thu May 11 06:50:22 2000

GN G. Turkey rhinotracheitis virus (TRTV).
 OS Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=2119;
 RX MEDLINE; 95053872.
 RA JUHASZ K., EASTON A.J., JUHASZ K., EASTON A.J.;
 RT "Extensive sequence variation in the attachment (G) protein gene of
 avian pneumovirus: evidence for two distinct subgroups.";
 RL J. Gen. Virol. 75:2873-2880(1994).
 DR EMBL; L34031; AAA68032.1;
 SQ SEQUENCE 414 AA; 44634 MW; BF2C72A3 CRC32;
 Query Match 81.5%; Score 44; DB 14; Length 414;
 Best Local Similarity 83.3%; Pred. No. 1.77e+01;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 116 MHDRSY 121
 QY 363 MNDRSY 368
 I:|||||

RESULT 15 PRELIMINARY; PRT; 435 AA.
 ID Q86788
 AC Q86788;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE GLYCOPROTEIN GG.
 OS Feline herpesvirus (Felid)herpesvirus 1).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirus.
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 94267406.
 RA SPATZ S.J., ROTA P.A., MAES R.K.;
 RT "Identification of the feline herpesvirus type 1 (FHV-1) genes
 encoding glycoproteins G, D, I and E: expression of FHV-1 glycoprotein
 D in vaccinia and raccoon poxviruses.";
 RL J. Gen. Virol. 75:1235-1244(1994).
 DR EMBL; S72415; AAB30979.1;
 SQ SEQUENCE 435 AA; 49112 MW; EB5534A0 CRC32;
 Query Match 81.5%; Score 44; DB 14; Length 435;
 Best Local Similarity 71.4%; Pred. No. 1.77e+01;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 409 MVDRAV 415
 QY 363 MNDRSY 369
 I:|||||

Search completed: Wed May 10 14:29:42 2000
 Job time : 236 secs.

QY 363 MNDRSY 368
 I:|||||

RESULT 12 PRELIMINARY; PRT; 414 AA.
 ID Q65306
 AC Q65306;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE ATTACHMENT PROTEIN.
 GN G. Turkey rhinotracheitis virus (TRTV).
 OS Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=5728;
 RX MEDLINE; 95053872.
 RA JUHASZ K., EASTON A.J., JUHASZ K., EASTON A.J.;
 RT "Extensive sequence variation in the attachment (G) protein gene of
 avian pneumovirus: evidence for two distinct subgroups.";
 RL J. Gen. Virol. 75:2873-2880(1994).
 DR EMBL; L34034; AAA68035.1;
 SQ SEQUENCE 414 AA; 44697 MW; 54487B12 CRC32;
 Query Match 81.5%; Score 44; DB 14; Length 414;
 Best Local Similarity 83.3%; Pred. No. 1.77e+01;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 116 MHDRSY 121
 QY 363 MNDRSY 368
 I:|||||

RESULT 13 PRELIMINARY; PRT; 414 AA.
 ID Q65305
 AC Q65305;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE ATTACHMENT PROTEIN.
 GN G. Turkey rhinotracheitis virus (TRTV).
 OS Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=6574;
 RX MEDLINE; 95053872.
 RA JUHASZ K., EASTON A.J., JUHASZ K., EASTON A.J.;
 RT "Extensive sequence variation in the attachment (G) protein gene of
 avian pneumovirus: evidence for two distinct subgroups.";
 RL J. Gen. Virol. 75:2873-2880(1994).
 DR EMBL; L34033; AAA68034.1;
 SQ SEQUENCE 414 AA; 44632 MW; 13DCF02E CRC32;
 Query Match 81.5%; Score 44; DB 14; Length 414;
 Best Local Similarity 83.3%; Pred. No. 1.77e+01;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

414 AA.

Sequence update)
 (notation update)

ted
se
imp

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 14:05:36 2000; MasPar time 3.00 Seconds
Tabular output not generated. 110.572 Million cell updates/sec

Title: >US-09-376-430-2
Description: (300-313) from US09376430A.pep (22 of 25)
Perfect Score: 88
Sequence: 1 AGAEGSGPEPLV 14

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a:geneseq35
1:geneseq

Statistics: Mean 16.761; Variance 52.643; scale 0.318

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	61	69.3	17	1 R07706	Rb-8 antigen derived f	8.40e+00
2	61	69.3	797	1 W69373	Modified retinoblastom	8.40e+00
3	61	69.3	859	1 W69372	Modified retinoblastom	8.40e+00
4	61	69.3	869	1 W69374	Modified retinoblastom	8.40e+00
5	61	69.3	871	1 W69375	Modified retinoblastom	8.40e+00
6	61	69.3	897	1 W69371	Modified retinoblastom	8.40e+00
7	61	69.3	928	1 Y01543	Human retinoblastoma p	8.40e+00
8	61	69.3	928	1 R74271	Retinoblastoma tumour	8.40e+00
9	61	69.3	928	1 W69376	Modified retinoblastom	8.40e+00
10	61	69.3	928	1 W71354	Protein sequence of th	8.40e+00
11	61	69.3	928	1 R71880	Retinoblastoma p110RB	8.40e+00
12	61	69.3	928	1 R06289	Predicted retinoblasto	8.40e+00
13	61	69.3	928	1 W09411	Retinoblastoma suscep	8.40e+00
14	61	69.3	928	1 R05305	Cancer suppressing gene	8.40e+00
15	61	69.3	928	1 R71681	Recombinant p110RB pr	8.40e+00
16	61	69.3	928	1 W69364	Modified retinoblastom	8.40e+00
17	61	69.3	928	1 R36334	Retinoblastoma (RB) pr	8.40e+00
18	61	69.3	928	1 W67801	Retinoblastoma protein	8.40e+00
19	61	69.3	928	1 W62465	Retinoblastoma protein	8.40e+00
20	61	69.3	928	1 W67802	Retinoblastoma protein	8.40e+00
21	55	62.5	392	1 R12345	Toxoplasma gondii prot	3.72e+01
22	55	62.5	428	1 R12352	Toxoplasma gondii p66	3.72e+01
23	54	61.4	895	1 W69365	Modified retinoblastom	4.75e+01

24	53	60.2	230	1	W09111	Human criptin growth f	6.06e+01
25	51	58.0	732	1	W72061	HSV-2 strain SB5 Conti	9.80e+01
26	51	58.0	785	1	W72103	HSV-2 strain SB5 Conti	9.80e+01
27	51	58.0	3119	1	W72204	HSV-2 strain SB5 Conti	9.80e+01
28	49	55.7	475	1	W94306	Thermomyces lanuginosu	1.58e+02
29	49	55.7	475	1	W35724	T. lanuginosus CBS 586	1.58e+02
30	49	55.7	475	1	W27384	Thermomyces lanuginosu	1.58e+02
31	49	55.7	710	1	W94509	Mouse OPF #5 protein.	1.58e+02
32	49	55.7	911	1	W90263	A. tigrinum AE1 protei	1.58e+02
33	49	55.7	911	1	R15355	Human erythrocyte memb	1.58e+02
34	49	55.7	2115	1	W59276	Rubella virus RA27/3 N	1.58e+02
35	49	55.7	2205	1	R79048	Infectious rubella vir	1.58e+02
36	48	54.5	147	1	R79649	C. albicans caUCE.	2.00e+02
37	48	54.5	148	1	R99264	Cotton plant ubiquitin	2.00e+02
38	48	54.5	180	1	W69601	Human zinc binding pro	2.00e+02
39	48	54.5	2647	1	W19349	Human filamin.	2.00e+02
40	47	53.4	147	1	R79650	S. pombe spUBCE.	2.52e+02
41	47	53.4	150	1	W00363	Human ubiquitin-conjug	2.52e+02
42	47	53.4	204	1	R95168	bcl-x(L)/bcl-2 associa	2.52e+02
43	47	53.4	204	1	W61315	Murine BCL-XL/BCL-2 as	2.52e+02
44	47	53.4	204	1	W58832	Murine BAD protein.	2.52e+02
45	47	53.4	473	1	W09874	Romaine lettuce violax	2.52e+02

ALIGNMENTS

RESULT 1
ID R07706 standard; peptide; 17 AA.

AC R07706;
DT 14-FEB-1991 (first entry)
DE Rb-8 antigen derived from retinoblastoma gene product.
KW retinoblastoma; Rb-8; carcinoma; diagnosis.
OS Synthetic.
PN W09012807-A.

PD 01-NOV-1990.
PF 13-APR-1990; U02017.
PR 14-APR-1989; US-338289.

PA (BAYU) BAYLOR COLLEGE MED.
PI Xu HJ, Benedict WF, Hu SX;
DR WPI; 90-348425/46.

PT New synthetic polypeptide antigens - used to produce highly specific antibodies against gene products, specifically retinoblastoma, for identifying and diagnosing carcinomas
PS Claim 21; Page 35; 52pp; English.
CC One of 7 antigenic peptides derived from the retinoblastoma (Rb) gene product. Rb-8 corresponds to amino acids 30-46 of the complete retinoblastoma protein sequence. The high hydrophobicity and high solubility of the peptide leads to the production of highly specific antibodies which can be used for determining prediagnosis to certain cancers. See also R07707-8 and R07710-R07712.

SQ Sequence 17 AA;

Query Match 69.3%; Score 61; DB 1; Length 17;
Best Local Similarity 69.2%; Pred. No. 8.40e+00;
Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 4 PEODSGPEPLV 16

Qy 302 AEQESGPEE-PLV 313

RESULT 2

ID W69373 standard; Protein; 797 AA.

AC W69373;
DT 02-DEC-1998 (first entry)
DE Modified retinoblastoma tumour suppressor.
KW Modified retinoblastoma tumour suppressor; RTSP protein; cancer therapy;
OS Cellular proliferation inhibitor.

PN W09837091-A2.

PD 27-AUG-1998.

PF 19-FEB-1998; U03041.

PR 20-FEB-1997; US-038118.

PA (BAYU) BAYLOR COLLEGE MEDICINE.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Benedict WF, Hu S, Xu H, Zhou Y;
 DR WPI: 98-480788/41.
 DR N-PSDB: V58449.
 PT Retinoblastoma suppressor protein with N-terminal modification -
 PT inhibiting cellular proliferation, particularly cancer
 PS Claim 21: Page 203-207; 249pp; English.
 CC This sequence represents a modified retinoblastoma tumour suppressor
 CC protein (RTSP) of the invention. The proteins can be used for inhibiting
 CC cellular proliferation, when coadministered with a p53 protein. The RTSPs
 CC can be used for treating diseases characterised by abnormal cellular
 CC proliferation, particularly cancers. The RTSPs have a broader spectrum of
 CC activity than wild type RTSPs.
 SQ Sequence 797 AA;

Query Match 69.3%; Score 61; DB 1; Length 797;

Best Local Similarity 69.2%; Pred. No. 8.40e+00;

Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEQDSGPEPLV 45

:||:||||: |||

QY 302 AEQSGPEE-PLV 313

RESULT 3

ID W69372 standard; Protein; 859 AA.

AC W69372;

DT 02-DEC-1998 (first entry)

DE Modified retinoblastoma tumour suppressor.

KW Modified retinoblastoma tumour suppressor; RTSP protein; cancer therapy;

KW cellular proliferation inhibitor.

OS Homo sapiens.

PN W09837091-A2.

PD 27-AUG-1998.

PF 19-FEB-1998; U03041.

PR 20-FEB-1997; US-038118.

PA (BAYU) BAYLOR COLLEGE MEDICINE.

PA (TEXA) UNIV TEXAS SYSTEM.

PI Benedict WF, Hu S, Xu H, Zhou Y;

DR WPI: 98-480788/41.

DR N-PSDB: V8448.

PT Retinoblastoma suppressor protein with N-terminal modification -

PT inhibiting cellular proliferation, particularly cancer

PS Claim 21: Page 196-200; 249pp; English.

CC This sequence represents a modified retinoblastoma tumour suppressor

CC protein (RTSP) of the invention. The proteins can be used for inhibiting

CC cellular proliferation, when coadministered with a p53 protein. The RTSPs

CC can be used for treating diseases characterised by abnormal cellular

CC proliferation, particularly cancers. The RTSPs have a broader spectrum of

CC activity than wild type RTSPs.

SQ Sequence 859 AA;

Query Match 69.3%; Score 61; DB 1; Length 859;

Best Local Similarity 69.2%; Pred. No. 8.40e+00;

Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEQDSGPEPLV 45

:||:||||: |||

QY 302 AEQSGPEE-PLV 313

RESULT 4

ID W69374 standard; Protein; 869 AA.

AC W69374;

DT 02-DEC-1998 (first entry)

DE Modified retinoblastoma tumour suppressor.

KW Modified retinoblastoma tumour suppressor; RTSP protein; cancer therapy;

KW cellular proliferation inhibitor.

OS Homo sapiens.

PN W09837091-A2.

PD 27-AUG-1998.

PF 19-FEB-1998; U03041.

PR 20-FEB-1997; US-038118.
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Benedict WF, Hu S, Xu H, Zhou Y;
 DR WPI: 98-480788/41.
 DR N-PSDB: V58450.
 PT Retinoblastoma suppressor protein with N-terminal modification -
 PT inhibiting cellular proliferation, particularly cancer
 PS Claim 21: Page 210-215; 249pp; English.
 CC This sequence represents a modified retinoblastoma tumour suppressor
 CC protein (RTSP) of the invention. The proteins can be used for inhibiting
 CC cellular proliferation, when coadministered with a p53 protein. The RTSPs
 CC can be used for treating diseases characterised by abnormal cellular
 CC proliferation, particularly cancers. The RTSPs have a broader spectrum of
 CC activity than wild type RTSPs.
 SQ Sequence 869 AA;

Query Match 69.3%; Score 61; DB 1; Length 869;

Best Local Similarity 69.2%; Pred. No. 8.40e+00;

Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEQDSGPEPLV 45

:||:||||: |||

QY 302 AEQSGPEE-PLV 313

RESULT 5

ID W69375 standard; Protein; 871 AA.

AC W69375;

DT 02-DEC-1998 (first entry)

DE Modified retinoblastoma tumour suppressor.

KW Modified retinoblastoma tumour suppressor; RTSP protein; cancer therapy;

KW cellular proliferation inhibitor.

OS Homo sapiens.

PN W09837091-A2.

PD 27-AUG-1998.

PF 19-FEB-1998; U03041.

PR 20-FEB-1997; US-038118.

PA (BAYU) BAYLOR COLLEGE MEDICINE.

PA (TEXA) UNIV TEXAS SYSTEM.

PI Benedict WF, Hu S, Xu H, Zhou Y;

DR WPI: 98-480788/41.

DR N-PSDB: V58451.

PT Retinoblastoma suppressor protein with N-terminal modification -

PT inhibiting cellular proliferation, particularly cancer

PS Claim 21: Page 218-222; 249pp; English.

CC This sequence represents a modified retinoblastoma tumour suppressor

CC protein (RTSP) of the invention. The proteins can be used for inhibiting

CC cellular proliferation, when coadministered with a p53 protein. The RTSPs

CC can be used for treating diseases characterised by abnormal cellular

CC proliferation, particularly cancers. The RTSPs have a broader spectrum of

CC activity than wild type RTSPs.

SQ Sequence 871 AA;

Query Match 69.3%; Score 61; DB 1; Length 871;

Best Local Similarity 69.2%; Pred. No. 8.40e+00;

Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEQDSGPEPLV 45

:||:||||: |||

QY 302 AEQSGPEE-PLV 313

RESULT 6

ID W69371 standard; Protein; 897 AA.

AC W69371;

DT 02-DEC-1998 (first entry)

DE Modified retinoblastoma tumour suppressor.

KW Modified retinoblastoma tumour suppressor; RTSP protein; cancer therapy;

KW cellular proliferation inhibitor.

OS Homo sapiens.

PN W09837091-A2.

PD 27-AUG-1998.

PF 19-FEB-1998; U03041.
 PR 20-FEB-1997; US-038118.
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Benedict WF, Hu S, Xu H, Zhou Y;
 DR WPI: 98-480788/41.
 DR N-PSDB; V58447.
 PT Retinoblastoma suppressor protein with N-terminal modification -
 PT inhibiting cellular proliferation, particularly cancer
 PS Claim 21; Page 188-192; 249pp; English.
 CC This sequence represents a modified retinoblastoma tumour suppressor
 CC protein (RTSP) of the invention. The proteins can be used for inhibiting
 CC cellular proliferation, when coadministered with a p53 protein. The RTSPs
 CC can be used for treating diseases characterised by abnormal cellular
 CC proliferation, particularly cancers. The RTSPs have a broader spectrum of
 CC activity than wild type RTSPs.
 SQ Sequence 897 AA;

Query Match 69.3%; Score 61; DB 1; Length 987;
 Best Local Similarity 69.2%; Pred. No. 8.40e+00;
 Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEQDSGPEDPLV 45

QY 302 AEQESGPEE-PLV 313

RESULT 7

ID Y01543 standard; Protein; 928 AA.
 AC Y01543;
 DT 05-JUL-1999 (first entry)
 DE Human retinoblastoma polypeptide.
 KW Retinoblastoma; Rb gene; human; predisposition; diagnosis; therapy.
 OS Homo sapiens.
 PN US5853988-A.
 PD 29-DEC-1998.
 PF 08-OCT-1992; 958290.
 PR 08-JUL-1991; US-728756.
 PR 11-AUG-1986; US-895163.
 PR 21-JAN-1988; US-146525.
 PR 23-JAN-1988; US-300667.
 PR 25-SEP-1992; US-951342.
 PR 08-OCT-1992; US-958290.
 PA (MASS-) MASSACHUSETTS EYE & EAR INFIRMARY.
 PA (WHEE) WHITEHEAD INST.
 PI Dryja TP, Friend S, Wandell DW;
 DR WPI: 99-094898/08.
 DR N-PSDB; X04502.
 PT Isolated human retinoblastoma gene - used to develop products for
 PT analysing predisposition to retinoblastoma and for the detection and
 PT treatment of retinoblastoma
 PS Disclosure; Fig 6; 32pp; English.
 CC This amino acid sequence was deduced from the nucleotide sequence
 CC of the normal human retinoblastoma gene (Rb) gene (see X04502).
 CC Absence or mutation of the Rb gene (see also X04502) predisposes an
 CC individual to retinoblastoma. Rb nucleic acids comprising the Rb
 CC gene, cDNA or a fragment of the gene, can be used in a claimed
 CC method for detecting a deletion or point mutation in the Rb gene of
 CC a human patient. The gene can also be used to synthesize Rb
 CC polypeptide for use in the treatment of individuals determined to
 CC have a defective Rb allele and hence to be at risk of developing
 CC retinoblastoma.
 SQ Sequence 928 AA;

Query Match 69.3%; Score 61; DB 1; Length 928;
 Best Local Similarity 69.2%; Pred. No. 8.40e+00;
 Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEQDSGPEDPLV 45

QY 302 AEQESGPEE-PLV 313

RESULT 8

ID R74271 standard; Protein; 928 AA.
 AC R74271;
 DT 28-NOV-1995 (first entry)
 DE Retinoblastoma tumour suppressor protein.
 KW Recombinant; adenovirus; expression vector; TSG; small lung cancer;
 KW hepatocarcinoma; melanoma; retinoblastoma; sarcoma; sickle cell;
 KW anaemia; Tay-Sach's disease.
 OS Homo sapiens.
 PN WO9511984-A.
 PD 04-MAY-1995.
 PF 25-OCT-1994; U12235.
 PR 25-OCT-1993; US-142669.
 PR 19-MAY-1994; US-246007.
 PA (CANU-) CANU INC.
 PI Gregory RJ, Maneval DC, Wills KN;
 DR WPI: 95-178876/23.
 DR N-PSDB; Q90059.
 PT Adenoviral vector with deletion of viral protein IX contains
 PT foreign gene - esp. encoding tumour suppressor protein for gene
 PT therapy of tumours, reduces contamination by wild type adenovirus
 PS Disclosure; Fig 3; 92pp; English.
 CC The sequence is that of a retinoblastoma tumour protein. The gene
 CC encoding this protein may be used in a novel method involving a
 CC recombinant adenovirus expression vector to treat diseases
 CC associated with the absence of the Tsg or the presence of a mutated
 CC TSG, e.g. many forms of carcinoma, sickle cell anaemia or Tay-Sach's
 CC disease.
 CC See also R74272.
 SQ Sequence 928 AA;

Query Match 69.3%; Score 61; DB 1; Length 928;
 Best Local Similarity 69.2%; Pred. No. 8.40e+00;
 Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEQDSGPEDPLV 45

QY 302 AEQESGPEE-PLV 313

RESULT 9

ID W69376 standard; Protein; 928 AA.
 AC W69376;
 DT 02-DEC-1998 (first entry)
 DE Modified retinoblastoma tumour suppressor.
 KW Modified retinoblastoma tumour suppressor; RTSP protein; cancer therapy;
 KW cellular proliferation inhibitor.
 OS Homo sapiens.
 PN WO9837091-A2.
 PD 27-AUG-1998.
 PF 19-FEB-1998; U03041.
 PR 20-FEB-1997; US-038118.
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Benedict WF, Hu S, Xu H, Zhou Y;
 DR WPI: 98-480788/41.
 DR N-PSDB; V58452.
 PT Retinoblastoma suppressor protein with N-terminal modification -
 PT inhibiting cellular proliferation, particularly cancer
 PS Claim 21; Page 226-230; 249pp; English.
 CC This sequence represents a modified retinoblastoma tumour suppressor
 CC protein (RTSP) of the invention. The proteins can be used for inhibiting
 CC cellular proliferation, when coadministered with a p53 protein. The RTSPs
 CC can be used for treating diseases characterised by abnormal cellular
 CC proliferation, particularly cancers. The RTSPs have a broader spectrum of
 CC activity than wild type RTSPs.
 SQ Sequence 928 AA;

Query Match 69.3%; Score 61; DB 1; Length 928;
 Best Local Similarity 69.2%; Pred. No. 8.40e+00;
 Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEQDSGPEDPLV 45

QY 302 AEQESGPEE-PLV 313
 :||:||||: |||

RESULT 10

ID W71354 standard; Protein; 928 AA.
 AC W71354, 1998 (first entry)
 DE Protein sequence of the specification.
 KW Minimal promoter; tetracycline responsive expression vector; TREV;
 KW transcriptional transactivation domain; TTD;
 KW tetracycline repressor protein; TRP; tetracycline operator;
 KW production; stable cell line; protein production;
 KW tumour suppressor protein; treatment; cancer.
 OS Unidentified.
 PN WO9837185-A2.
 PD 27-AUG-1998.
 PF 19-FEB-1998; U03092.
 PR 20-FEB-1997; US-038755.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Hu S, Logothetis CJ, Xu H, Zhou Y;
 DR WPI; 98-480796/41.
 DR N-PSDB; V54990.
 PT New tetracycline responsive expression vectors - used for the
 PT tightly controlled expression of genes, such as tumour suppressor
 PT genes for treating cancers
 PS Disclosure; Pages 150-153; 190pp; English.
 CC The present sequence appears in the specification, which describes a
 CC tetracycline responsive expression vector (TREV), which contains a
 CC sequence encoding a fusion protein comprising a transcriptional
 CC transactivation domain (TTD) operatively attached to a tetracycline
 CC repressor protein (TRP). The first sequence operatively positioned
 CC downstream of a promoter. The vector also contains a cloning site
 CC operatively positioned downstream of a basal promoter comprising a
 CC tetracycline operator. The TREVs can be used to produce stable cell
 CC lines in which gene expression is tightly regulated by tetracycline.
 CC They can be used for the production of proteins such as tumour
 CC suppressor proteins which can be used for treating diseases characterised
 CC by abnormal cellular proliferation, particularly cancers.
 SQ Sequence 928 AA;

Query Match 69.3%; Score 61; DB 1; Length 928;
 Best Local Similarity 69.2%; Pred. No. 8.40e+00;
 Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEQDSGPEDLPLV 45
 :||:||||: |||

RESULT 11

ID R71680 standard; Protein; 928 AA.
 AC R71680;
 DT 16-OCT-1995 (first entry)
 DE Retinoblastoma pRB10RB protein.
 KW Rb10; retinoblastoma; cancer; tumor suppressor protein;
 KW cell cycle; pRB10RB.
 OS Homo sapiens.
 PN WO9507708-A.
 PD 23-MAR-1995.
 PF 13-SEP-1994; U10357.
 PR 13-SEP-1993; US-121108.
 PA (CANJ-) CANJI INC.
 PA (REGC) UNIV CALIFORNIA.
 PI Goodrich DW, Johnson D, Lee EXP, Lee W, Shepard HM;
 PI Wang NP;
 DR WPI; 95-131179/17.
 DR N-PSDB; Q86398.
 PT Admin. of a functional retinoblastoma polypeptide or protein -
 PT used to prevent and inhibit prim. and sec. retinoblastoma-linked
 PT cancers
 PS Disclosure; Fig.2; 163pp; English.
 CC By chromosomal walking from the esterase D gene on chromosome 13,

CC the retinoblastoma susceptibility gene (RB) was identified on the
 CC basis of chromosomal location, homologous deletion and tumor-specific
 CC alterations in expression. Screening of cDNA libraries yielded clone
 CC RB-5 incorporating the complete RB cDNA sequence given in Q86398 and
 CC encoding a 110-140 kba nuclear phosphoprotein named pRB10RB (R71680).
 CC Use of recombinant pRB10RB will reduce the need for conventional
 CC radiotherapy or chemotherapy.
 SQ Sequence 928 AA;

Query Match 69.3%; Score 61; DB 1; Length 928;
 Best Local Similarity 69.2%; Pred. No. 8.40e+00;
 Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEQDSGPEDLPLV 45
 :||:||||: |||

QY 302 AEQESGPEE-PLV 313

RESULT 12

ID R06289 standard; protein; 928 AA.
 AC R06289;
 DT 13-DEC-1990 (first entry)
 DE Predicted retinoblastoma gene product.
 KW Osteosarcoma; fibrosarcoma; glioblastoma; breast cancer; ds.
 OS Homo sapiens.
 PN US4942123-A.
 PD 17-JUL-1990.
 PF 17-SEP-1987; 098612.
 PR 17-SEP-1987; US-098612.
 PA (REGC) UNIV OF CALIFORNIA.
 PI Lee WH, Eva Y, Lee HP;
 DR WPI; 90-245977/32.
 PT Diagnosing absence or inactivation of retinoblastoma gene - by
 PT detecting the absence of specific anti-pRB 110 antibody
 PT immuno-complex formed using tissue
 PS Disclosure; P; English.
 CC Laelled Abs raised to the RB gene product may be used to screen
 CC for RB and in diagnosis of susceptibility to associated secondary
 CC cancers such as osteosarcoma, fibrosarcoma, glioblastoma and
 CC breast cancer.
 SQ Sequence 928 AA;

Query Match 69.3%; Score 61; DB 1; Length 928;
 Best Local Similarity 69.2%; Pred. No. 8.40e+00;
 Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEQDSGPEDLPLV 45
 :||:||||: |||

QY 302 AEQESGPEE-PLV 313

RESULT 13

ID W09411 standard; Protein; 928 AA.
 AC W09411;
 DT 22-JUN-1997 (first entry)
 DE Retinoblastoma susceptibility phosphoprotein pRB110.
 KW Retinoblastoma susceptibility gene; pRB110; nuclear phosphoprotein;
 KW cancer; osteosarcoma; fibrosarcoma; glioblastoma; breast cancer;
 KW polyclonal antibody; diagnosis.
 OS Homo sapiens.
 PN US5578701-A.
 PD 26-NOV-1996.
 PF 17-SEP-1987; 098612.
 PR 17-SEP-1987; US-098612.
 PR 11-JUL-1990; US-550877.
 PR 14-JUL-1992; US-914039.
 PR 17-JUN-1993; US-079207.
 PR 08-APR-1994; US-225099.
 PA (REGC) UNIV CALIFORNIA.
 PI Lee EXP, Lee W;
 DR WPI; 97-020465/02.
 PT Retinoblastoma phosphoprotein pRB110-specific polyclonal antibody
 PT - for diagnosing retinoblastoma and other related tumours or

PT susceptibility to them
 PS Disclosure; Fig 2; 21pp; English.
 CC A 110-114 kDa phosphoprotein product (W09411) of the retinoblastoma
 CC susceptibility gene is designated pRB110. Its sequence was deduced
 CC from a full-length cDNA sequence that included the sequence of a
 CC clone derived from the retinoblastoma cell line Y79. pRB110 is
 CC primarily located in the cell nucleus and has DNA binding activity.
 CC The absence or loss of pRB10 mediates oncogenicity. pRB110-
 CC specific polyclonal antibodies can be utilised in methods of
 CC diagnosing hereditary predisposition to retinoblastoma or to other
 CC diseases controlled by the retinoblastoma gene such as
 CC osteosarcoma, fibrosarcoma, glioblastoma and breast cancer.
 CC Provision of pRB110 to an individual through molecular induction
 CC and gene transplanting may be used as a means of suppressing
 CC tumorigenesis.
 SQ Sequence 928 AA;

Query Match 69.3%; Score 61; DB 1; Length 928;
 Best Local Similarity 69.2%; Pred. No. 8.40e+00;
 Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEQDSGPEPLPLV 45
 :||:||||: |||
 QY 302 AEQESGPEE-PLV 313

RESULT 14
 ID R05305 standard; protein; 928 AA.
 AC R05305;
 DT 11-OCT-1990 (first entry)
 DE Cancer suppressing gene (CSG) product.
 KW Cancer; cancer suppressing gene; CSG; 13q14; retinoblastoma;
 KW RB; ds.
 OS Homo sapiens.
 PN W09005180-A.
 PD 17-MAY-1990.
 PF 30-OCT-1989; 004808.
 PR 31-OCT-1988; US-265829.
 PA (REGC) Univ of California.
 PI Lee WH, Huang HJS;
 DR WPI; 90-178822/23.
 DR N-PSDB; Q04713.
 PT Controlling cancer -
 PT by replacing ineffective cancer suppressing gene with cloned,
 PT active gene.
 PS Claim 35; Page 86; 105pp; English.
 CC Gene is taken from human chromosome 13q14 retinoblastoma (RB) cDNA.
 CC By installing a working CSG, safe and specific treatment and
 CC prophylaxis can be given to cancer patients.
 SQ Sequence 928 AA;

Query Match 69.3%; Score 61; DB 1; Length 928;
 Best Local Similarity 69.2%; Pred. No. 8.40e+00;
 Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEQDSGPEPLPLV 45
 :||:||||: |||
 QY 302 AEQESGPEE-PLV 313

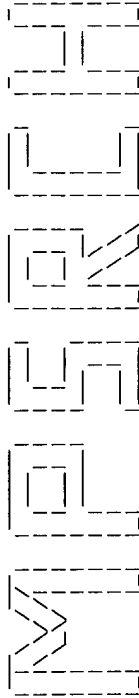
RESULT 15
 ID R71681 standard; Protein; 928 AA.
 AC R71681;
 DT 16-OCT-1995 (first entry)
 DE Recombinant pR10RB protein.
 KW Rb110; retinoblastoma; cancer; tumor suppressor protein;
 KW cell cycle; pR10RB; Escherichia coli.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_difference 2
 FT /note= "amino acid at position 2 is Pro in the
 FT natural sequence"
 PN W09507708-A.

PD 23-MAR-1995.
 PF 13-SEP-1994; U10357.
 PR 13-SEP-1993; US-121108.
 PA (CANJ-) CANJ INC.
 PA (REGC) UNIV CALIFORNIA.
 PI Goodrich DW, Johnson D, Lee EYP, Lee W, Shepard HM;
 PI Wang NP;
 DR WPI; 95-131179/17.
 PT Admin. of a functional retinoblastoma polypeptide or protein -
 PT used to prevent and inhibit prim. and sec. retinoblastoma-linked
 PT cancers
 PS Disclosure; Fig.30; 163pp; English.
 CC By chromosomal walking from the esterase D gene on chromosome 13,
 CC the retinoblastoma susceptibility gene (RB) was identified on the
 CC basis of chromosomal location, homologous deletion and tumor-specific
 CC alterations in expression. Screening of cDNA libraries yielded clone
 CC RB-5 incorporating the complete RB cDNA sequence given in Q86398 and
 CC encoding a 110-140 kDa nuclear phosphoprotein named pR10RB (R71680).
 CC Recombinant pR10RB was prepared in E. coli, but with a P2A mutation
 CC for cloning convenience (R71681).
 SQ Sequence 928 AA;

Query Match 69.3%; Score 61; DB 1; Length 928;
 Best Local Similarity 69.2%; Pred. No. 8.40e+00;
 Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEQDSGPEPLPLV 45
 :||:||||: |||
 QY 302 AEQESGPEE-PLV 313

Search completed: Wed May 10 14:05:43 2000
 Job time : 7 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 14:05:59 2000; Maspar time 46.66 Seconds
Tabular output not generated. 3.888 Million cell updates/sec

Title: >US-09-376-430-2
Description: (300-313) from US09376430A.pep (22 of 25)
Perfect Score: 88
Sequence: 1 AGAEQESGPPEPLV 14

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 15.915; Variance 50.384; scale 0.316

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	61	69.3	928	1	US-08-204-Sequence 1, Applicatio	4.11e+00
2	61	69.3	928	1	PCT-US94-1Sequence 3, Applicatio	4.11e+00
3	61	69.3	928	2	US-08-482-Sequence 5, Applicatio	4.11e+00
4	61	69.3	928	3	PCT-US94-1Sequence 2, Applicatio	4.11e+00
5	61	69.3	928	2	US-08-959-Sequence 8, Applicatio	4.11e+00
6	53	60.2	230	2	US-08-471-Sequence 2, Applicatio	3.03e+01
7	49	55.7	475	2	US-08-819-Sequence 2, Applicatio	7.98e+01
8	49	55.7	2205	1	US-08-093-Sequence 2, Applicatio	1.01e+02
9	48	54.5	222	1	US-08-080-Sequence 3, Applicatio	1.01e+02
10	48	54.5	147	2	US-08-486-Sequence 4, Applicatio	1.01e+02
11	48	54.5	147	1	US-08-305-Sequence 4, Applicatio	1.01e+02
12	48	54.5	148	2	US-08-464-Sequence 9, Applicatio	1.01e+02
13	48	54.5	148	2	US-09-196-Sequence 4, Applicatio	1.01e+02
14	48	54.5	148	2	US-08-679-Sequence 4, Applicatio	1.01e+02
15	48	54.5	180	2	US-08-786-Sequence 3, Applicatio	1.01e+02
16	48	54.5	180	2	US-08-933-Sequence 8, Applicatio	1.01e+02
17	48	54.5	2647	2	US-08-779-Sequence 8, Applicatio	1.01e+02
18	48	54.5	2647	2	US-08-583-Sequence 34, Applicatio	1.01e+02
19	47	53.4	22	1	US-08-080-Sequence 8, Applicatio	1.28e+02
20	47	53.4	147	2	US-08-247-Sequence 2, Applicatio	1.28e+02
21	47	53.4	147	2	US-08-464-Sequence 2, Applicatio	1.28e+02
22	47	53.4	147	2	US-08-748-Sequence 4, Applicatio	1.28e+02
23	47	53.4	147	2	US-08-903-Sequence 2, Applicatio	1.28e+02

24	47	53.4	147	2	US-08-464-Sequence 7, Applicatio	1.28e+02
25	47	53.4	147	2	US-08-989-Sequence 3, Applicatio	1.28e+02
26	47	53.4	147	2	US-08-486-Sequence 6, Applicatio	1.28e+02
27	47	53.4	147	2	US-08-486-Sequence 2, Applicatio	1.28e+02
28	47	53.4	147	2	US-08-875-Sequence 2, Applicatio	1.28e+02
29	47	53.4	147	1	US-08-305-Sequence 2, Applicatio	1.28e+02
30	47	53.4	147	1	US-08-464-Sequence 2, Applicatio	1.28e+02
31	47	53.4	147	1	US-08-305-Sequence 6, Applicatio	1.28e+02
32	47	53.4	204	2	US-08-733-Sequence 1, Applicatio	1.28e+02
33	47	53.4	204	2	US-08-717-Sequence 3, Applicatio	1.28e+02
34	47	53.4	204	2	US-08-733-Sequence 14, Applicatio	1.28e+02
35	47	53.4	204	2	US-08-733-Sequence 12, Applicatio	1.28e+02
36	47	53.4	204	2	US-08-733-Sequence 13, Applicatio	1.28e+02
37	47	53.4	204	2	US-08-661-Sequence 2, Applicatio	1.28e+02
38	47	53.4	204	1	US-08-333-Sequence 2, Applicatio	1.28e+02
39	47	53.4	373	1	US-08-625-Sequence 2, Applicatio	1.28e+02
40	47	53.4	617	1	US-08-340-Sequence 3, Applicatio	1.28e+02
41	47	53.4	617	3	PCT-US92-0Sequence 3, Applicatio	1.28e+02
42	47	53.4	617	1	US-08-753-Sequence 11, Applicatio	1.28e+02
43	47	53.4	617	1	US-08-301-Sequence 3, Applicatio	1.28e+02
44	47	53.4	617	1	US-07-879-Sequence 11, Applicatio	1.28e+02
45	46	52.3	477	2	US-08-452-Sequence 25, Applicatio	1.63e+02

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	928 AA.
ID	US-08-204-329-1			
XX	XXXXXX			
AC				
XX				
DT				
XX				
DE	Sequence 1, Application US/08204329			
CC	Sequence 1, Application US/08204329			
CC	Patent No. 5710255			
CC	GENERAL INFORMATION:			
CC	APPLICANT: SHEPARD, H. M.			
CC	APPLICANT: WEN, SHU F.			
CC	TITLE OF INVENTION: CHARACTERIZATION OF A NOVEL ANTI-P110RB			
CC	TITLE OF INVENTION: MONOCLONAL ANTIBODY			
CC	NUMBER OF SEQUENCES: 2			
CC	CORRESPONDENCE ADDRESS:			
CC	ADDRESSEE: TOWNSEND & TOWNSEND & CREW LLP			
CC	STREET: TWO EMBARCADERO CENTER, 8TH FLOOR			
CC	CITY: SAN FRANCISCO			
CC	STATE: CALIFORNIA			
CC	COUNTRY: U.S.A.			
CC	ZIP: 94111			
CC	COMPUTER READABLE FORM:			
CC	MEDIUM TYPE: Floppy disk			
CC	COMPUTER: IBM PC Compatible			
CC	OPERATING SYSTEM: PC-DOS/MS-DOS			
CC	SOFTWARE: Patent in Release #1.0, Version #1.30			
CC	CURRENT APPLICATION DATA:			
CC	APPLICATION NUMBER: US/08/204,329			
CC	FILING DATE: 15-AUG-1994			
CC	CLASSIFICATION: 435			
CC	PRIOR APPLICATION DATA:			
CC	APPLICATION NUMBER: PCT/US92/05866			
CC	FILING DATE: 14-JUL-1992			
CC	ATTORNEY/AGENT INFORMATION:			
CC	NAME: RENE A. FITTS			
CC	REGISTRATION NUMBER: 35,136			
CC	REFERENCE/DOCKET NUMBER: 16930-000400US			
CC	TELECOMMUNICATION INFORMATION:			
CC	TELEPHONE: (415) 326-2400			
CC	TELEPHONE: (415) 326-2422			
CC	INFORMATION FOR SEQ ID NO: 1:			
CC	SEQUENCE CHARACTERISTICS:			
CC	LENGTH: 928 amino acids			
CC	TYPE: amino acid			

CC	OPERATING SYSTEM:	PC-DOS/MS-DOS
CC	SOFTWARE:	Patentin Release #1.0, Version #1.25
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/08/959,638
CC	FILING DATE:	
CC	CLASSIFICATION:	
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/08/328,673
CC	FILING DATE:	25-OCT-1994
CC	APPLICATION NUMBER:	US 08/233,777
CC	FILING DATE:	19-MAY-1994
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	US 08/142,669
CC	FILING DATE:	25-OCT-1993
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	Campbell, Cathryn A.
CC	REGISTRATION NUMBER:	31,815
CC	REFERENCE/DOCKET NUMBER:	P-CJ 1192
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE:	(619) 535-9001
CC	TELEFAX:	(619) 535-8949
CC	INFORMATION FOR SEQ ID NO:	8:
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	928 amino acids
CC	TYPE:	amino acid
CC	TOPOLOGY:	linear
CC	MOLECULE TYPE:	protein
CC	SEQUENCE	928 AA; 106158 MW; 4596348 CN;
SQ		
Db	Query Match	69.3%; Score 61; DB 2: Length 9
	Best Local Similarity	69.2%; Pred. No. 4.11e+00;
	Matches	9; Conservative 3; Mismatches 0; Indels 0
QY		
Db	33 PEODSGPEDPLV 45	
	: : :	
QY	302 AEQSGPEE-PLV 313	
RESULT	6	
ID	US-08-471-371-2	STANDARD; PRT; 230 AA.
XX	xxxxxx	
AC		
XX		
DT		
XX		
DE		
XX		
CC	Sequence 2, Application US/08471371	
CC	Sequence 2, Application US/08471371	
CC	Patent No. 5981215	
CC	GENERAL INFORMATION:	
CC	APPLICANT:	Meissner, Paul S.
CC	APPLICANT:	Coleman, Timothy A.
CC	TITLE OF INVENTION:	HUMAN CRIPTIN GROWTH FACTOR
CC	NUMBER OF SEQUENCES:	7
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE:	CARELLA, BYRNE, BAIN, GILFILLAN, CECIL
CC	ADDRESS:	STEWART & OLSTEIN
CC	STREET:	6 Becker Farm Road
CC	CITY:	Roseland
CC	STATE:	N.J.
CC	COUNTRY:	U.S.A
CC	ZIP:	07068
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Floppy disk
CC	COMPUTER:	IBM PC compatible
CC	OPERATING SYSTEM:	PC-DOS/MS-DOS
CC	SOFTWARE:	Patentin Release #1.0, Version #1.30
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/08/471,371
CC	FILING DATE:	06-JUN-1995
CC	CLASSIFICATION:	435
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	Ferraro, Gregory D.

CC REGISTRATION NUMBER: 36,134
CC REFERENCE/DOCKET NUMBER: 32580-455
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (201) 994-1700
CC TELEFAX: (201) 994-1744
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 230 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 230 AA; 24919 MW; 243864 CN;

Query Match 60.2%; Score 53; DB 2; Length 230;
Best Local Similarity 53.8%; Pred. No. 3.03e+01;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 62 TGSAGWGPEEPL 74
QY 300 AGAEGSGPEEPL 312
:|: : |||||

RESULT 7
ID US-08-819-825-2 STANDARD; PRT; 475 AA.
XX AC xxxxxx
XX DT
XX DE
XX SE Sequence 2, Application US/08819825
XX Patent No. 5866118
XX GENERAL INFORMATION:
XX APPLICANT: Berkla, Randy M.
XX APPLICANT: Ray, Michael W.
XX APPLICANT: Klotz, Alan V.
XX TITLE OF INVENTION: Polypeptides Having Phytase Activity
XX TITLE OF INVENTION: And Nucleic Acids Encoding Same
XX NUMBER OF SEQUENCES: 5
XX CORRESPONDENCE ADDRESS:
XX ADDRESSEE: No. 58661180 No. 5866118disk of No. 5866118th America, Inc.
XX STREET: 405 Lexington Avenue, Suite 6400
XX CITY: New York
XX STATE: New York
XX COUNTRY: U.S.A.
XX ZIP: 10174-6401
XX COMPUTER READABLE FORM:
XX MEDIUM TYPE: Diskette
XX COMPUTER: IBM Compatible
XX OPERATING SYSTEM: DOS
XX SOFTWARE: Fast-Seq for Windows Version 2.0
XX CURRENT APPLICATION DATA:
XX APPLICATION NUMBER: US/08/819,825
XX FILING DATE: 18-MAR-1997
XX CLASSIFICATION: 435
XX ATTORNEY/AGENT INFORMATION:
XX NAME: Lambiris, Elias J.
XX REGISTRATION NUMBER: 33,728
XX REFERENCE/DOCKET NUMBER: 4758,200-US
XX TELECOMMUNICATION INFORMATION:
XX TELEPHONE: 212 867 0123
XX TELEFAX: 212 867 0298
XX INFORMATION FOR SEQ ID NO: 2:
XX SEQUENCE CHARACTERISTICS:
XX LENGTH: 475 amino acids
XX TYPE: amino acid
XX STRANDEDNESS: single
XX TOPOLOGY: linear
XX MOLECULE TYPE: protein
XX FRAGMENT TYPE: internal
SQ SEQUENCE 475 AA; 53278 MW; 1142990 CN;

Query Match 55.7%; Score 49; DB 2; Length 475;
Best Local Similarity 42.9%; Pred. No. 7.98e+01;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 418 SSEEEEGEPEFV 431
QY 300 AGAEGSGPEEPLV 313
:|:| :|:|:|

RESULT 8
ID US-08-093-453B-2 STANDARD; PRT; 2205 AA.
XX AC xxxxxx
XX DT
XX DE
XX SE Sequence 2, Application US/08093453B
XX Patent No. 5439814
XX GENERAL INFORMATION:
XX APPLICANT: Frey, Teryl K.
XX APPLICANT: Dominguez, Geraldina
XX APPLICANT: Wang, Chin-Yen
XX TITLE OF INVENTION: Modified Infectious Rubella Virus
XX NUMBER OF SEQUENCES: 13
XX CORRESPONDENCE ADDRESS:
XX ADDRESSEE: Jamie L. Greene, Jones & Askew
XX STREET: 191 Peachtree Street, 37th Floor
XX CITY: Atlanta
XX STATE: Georgia
XX COUNTRY: United States
XX ZIP: 30303
XX COMPUTER READABLE FORM:
XX MEDIUM TYPE: Diskette, 3.50
XX COMPUTER: Macintosh
XX OPERATING SYSTEM: 7.0
XX SOFTWARE: Microsoft Word
XX CURRENT APPLICATION DATA:
XX APPLICATION NUMBER: US/08/093,453B
XX FILING DATE: 19 JUL 1993
XX CLASSIFICATION: 435
XX PRIOR APPLICATION DATA:
XX APPLICATION NUMBER: U.S. 07/722,334
XX FILING DATE: 28 JUN 1991
XX ATTORNEY/AGENT INFORMATION:
XX NAME: Greene, Jamie L.
XX REGISTRATION NUMBER: 32,467
XX REFERENCE/DOCKET NUMBER: 07362-0101
XX TELECOMMUNICATION INFORMATION:
XX TELEPHONE: 404-818-3700
XX TELEFAX: 404-818-3799
XX INFORMATION FOR SEQ ID NO: 2:
XX SEQUENCE CHARACTERISTICS:
XX LENGTH: 2205 amino acids
XX TYPE: amino acid
XX STRANDEDNESS: single
XX TOPOLOGY: linear
XX MOLECULE TYPE: peptide
XX HYPOTHETICAL: NO
XX ANTI-SENSE: NO
XX FRAGMENT TYPE: N-terminal
XX ORIGINAL SOURCE:
XX ORGANISM: Rubella virus
XX STRAIN: Therien
SQ SEQUENCE 2205 AA; 240235 MW; 22500019 CN;

Query Match 55.7%; Score 49; DB 1; Length 2205;
Best Local Similarity 46.2%; Pred. No. 7.98e+01;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 1232 GSEHEAPPDHLV 1244
:|:|:| :|:|

Query Match 54.5%; Score 48; DB 2; Length 148;
Best Local Similarity 44.4%; Pred. No. 1.01e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 113 DANPDDPLV 121
::|:||||
QY 305 ESGPEEPLV 313

RESULT 14
ID US-08-679-765-4 STANDARD; PRT; 148 AA.
XX
AC xxxxxx
XX
DT

Sequence 4, Application US/08679765
Sequence 4, Application US/08679765
Patent No. 5840866
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Goli, Surya K.
APPLICANT: Hillman, Jennifer L.
TITLE OF INVENTION: A NOVEL HUMAN UBIQUITIN-CONJUGATING ENZYME
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/679,765
FILING DATE: Filed Herewith
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0093 US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 148 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 4718
SQ SEQUENCE 148 AA; 16456 MW; 118864 CN;

Query Match 54.5%; Score 48; DB 2; Length 148;
Best Local Similarity 44.4%; Pred. No. 1.01e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 113 DANPDDPLV 121
::|:||||
QY 305 ESGPEEPLV 313

RESULT 15
ID US-08-786-606-3 STANDARD; PRT; 180 AA.
XX
AC xxxxxx

XX
DI
XX

Sequence 3, Application US/08786606
Sequence 3, Application US/08786606
Patent No. 5861495

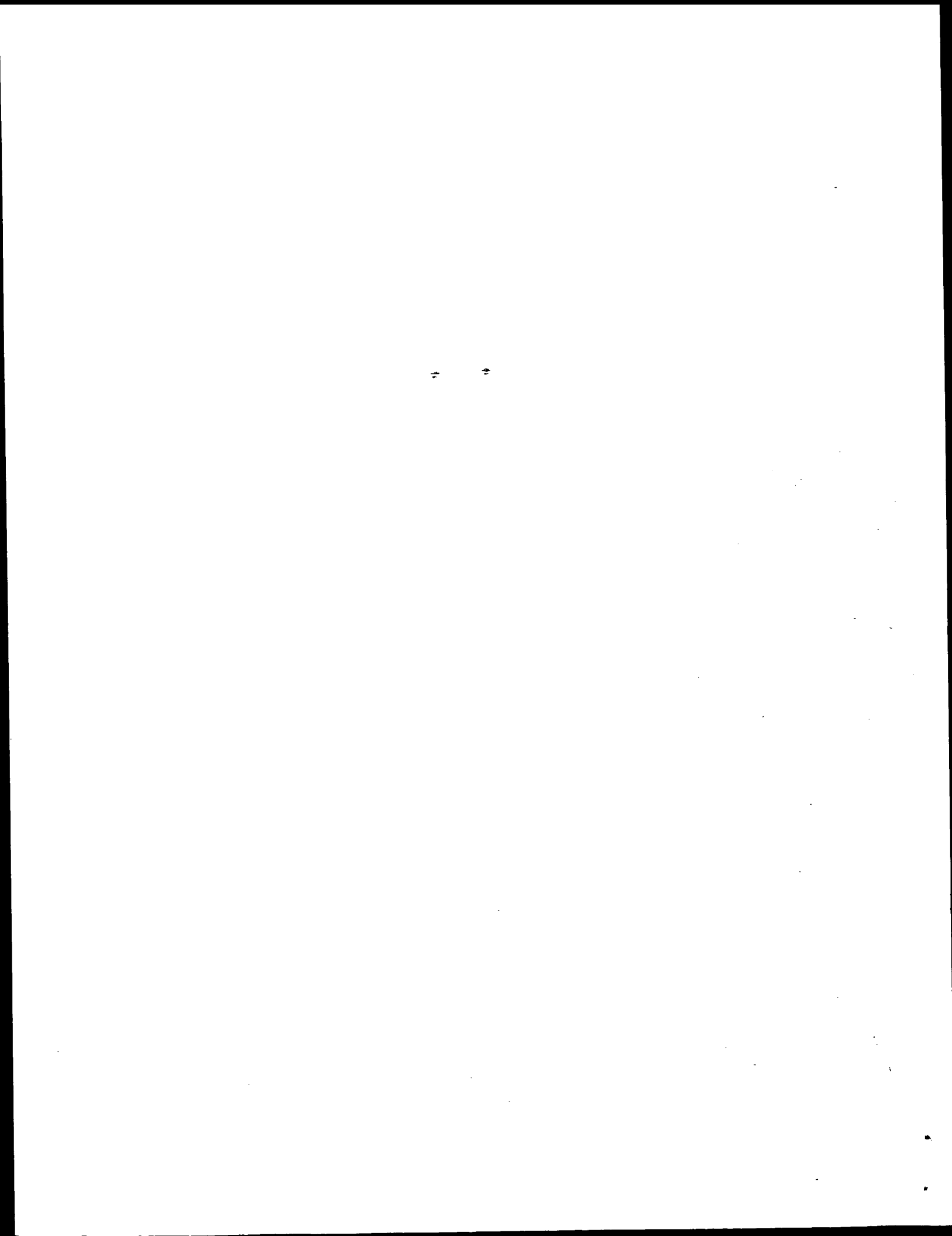
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Au-Young, Janice
APPLICANT: Coleman, Roger
APPLICANT: Goli, Surya K.
TITLE OF INVENTION: NOVEL HUMAN ZINC-BINDING
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/786,606
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy RJ
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0173 US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 180 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SQ SEQUENCE 180 AA; 19881 MW; 173541 CN;

Query Match 54.5%; Score 48; DB 2; Length 180;
Best Local Similarity 50.0%; Pred. No. 1.01e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Db 3 AAEEDGGPEGP 14
::|:||||
QY 300 AGAEQESGPEEP 311

Search completed: Wed May 10 14:06:55 2000
Job time : 56 secs.



W P L E H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 14:05:11 2000; MasPar time 4.25 Seconds
Tabular output not generated. 155.259 Million cell updates/sec

Title: >US-09-376-430-2
Description: (300-313) from US09376430A.pep (22 of 25)
Perfect Score: 88
Sequence: 1 AGAEQSGPEEPLV 14

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 23.443; Variance 30.959; scale 0.757

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	61	69.3	928	1 RBHU	retinoblastoma-associ	8.34e-02
2	56	63.6	1300	2 T03166	probable immediate ea	9.20e-01
3	55	62.5	396	1 A58938	surface protein rhopt	1.47e+00
4	53	60.2	88	2 S25078	acyl carrier protein	3.67e+00
5	53	60.2	317	2 C38178	Telb protein - plasm	3.67e+00
6	52	59.1	114	2 A30229	procytic acidic repe	5.75e+00
7	51	58.0	232	2 D70537	hypothetical protein	8.97e+00
8	51	58.0	356	1 A40701	tenascin-X precursor	8.97e+00
9	49	55.7	216	2 T02201	probable inorganic py	2.15e+01
10	49	55.7	243	2 C37772	ankyrin, erythrocyte	2.15e+01
11	49	55.7	316	2 I45996	apolipoprotein E - bo	2.15e+01
12	49	55.7	316	2 S26478	apolipoprotein E - bo	2.15e+01
13	49	55.7	525	2 D70747	probable fth protein	2.15e+01
14	49	55.7	911	1 B3RU	band 3 anion transpor	2.15e+01
15	49	55.7	921	2 A33718	retinoblastoma protei	2.15e+01
16	49	55.7	1862	2 I49502	ankyrin - mouse	2.15e+01
17	49	55.7	2115	2 S38480	nonstructural protein	2.15e+01
18	49	55.7	2205	1 MNWVRN	nonstructural polypro	2.15e+01
19	49	55.7	2588	2 T14342	NSDI protein - mouse	2.15e+01
20	48	54.5	148	2 S22858	ubiquitin--protein li	3.29e+01
21	48	54.5	148	2 S22858	ubiquitin--protein li	3.29e+01
22	48	54.5	180	2 T09063	hypothetical protein	3.29e+01
23	48	54.5	228	2 D69317	hypothetical protein	3.29e+01

24	48	54.5	278	1 TPHTW	troponin T, slow skel	3.29e+01
25	48	54.5	533	2 S33744	cyclodiene insecticid	3.29e+01
26	48	54.5	572	2 G02845	zyxin - human	3.29e+01
27	48	54.5	575	2 JG0181	XiLL2 protein - human	3.29e+01
28	48	54.5	867	2 A49284	immediate-early prote	3.29e+01
29	48	54.5	1221	2 A49457	fibulin-2 precursor -	3.29e+01
30	48	54.5	1873	2 A55845	calcium channel, volt	3.29e+01
31	48	54.5	2325	2 A61208	gelation factor ABP-2	3.29e+01
32	48	54.5	2647	2 A37098	chondroitin sulfate p	3.29e+01
33	48	54.5	3014	2 JC5620	polyprotein - hepatit	3.29e+01
34	47	53.4	118	2 S32673	ubiquitin--protein li	5.01e+01
35	47	53.4	147	2 S53358	ubiquitin conjugating	5.01e+01
36	47	53.4	147	2 A48145	ubiquitin conjugating	5.01e+01
37	47	53.4	147	2 I59365	ubiquitin conjugating	5.01e+01
38	47	53.4	147	2 I39202	ubiquitin conjugating	5.01e+01
39	47	53.4	204	2 A55671	bad protein - mouse	5.01e+01
40	47	53.4	314	2 S76315	hypothetical protein	5.01e+01
41	47	53.4	378	2 G02313	CDC37 homolog - human	5.01e+01
42	47	53.4	380	2 JL0053	UDPglucose--hexose-1-	5.01e+01
43	47	53.4	652	2 I40676	transcription initiat	5.01e+01
44	47	53.4	739	1 JQ1893	80.7K alpha trans-ind	5.01e+01
45	47	53.4	1265	2 S57968	Ran-binding protein 2	5.01e+01

ALIGNMENTS

RESULT 1

ENTRY	RBHU	#type complete
TITLE	retinoblastoma-associated	
ALTERNATE_NAMES	retinoblastoma susceptibility protein	
ORGANISM	#formal_name Homo sapiens #common_name man	
DATE	30-Jun-1987 #sequence_revision 30-Jun-1990 #text_change 22-Jun-1999	
ACCESSIONS	JS0276; A03152; A91613; A39947; A44987; I54364; I58362; I78863; I78866; I78872; I78873; A35590	
REFERENCE	JS0276	
#authors	Lee, W.H.; Shew, J.Y.; Hong, F.D.; Sery, T.W.; Donoso, L.A.; Young, L.J.; Bookstein, R.; Lee, E.Y.H.P.	
#journal	Nature (1987) 329:642-645	
#title	The retinoblastoma susceptibility gene encodes a nuclear phosphoprotein associated with DNA binding activity.	
#cross-references	MUID:88014238	
#accession	JS0276	
#molecule_type	mRNA	
#residues	1-928	
#cross-references	GB:M28419; NID:g190962; PIDN:AAA69808.1; PID:g190963	
#note	this sequence has two possible initiation sites, 1-Met and 113-Met	
REFERENCE	A03152	
#authors	Lee, W.H.; Bookstein, R.; Hong, F.; Young, L.J.; Shew, J.Y.; Lee, E.Y.H.P.	
#journal	Science (1987) 235:1394-1399	
#title	Human retinoblastoma susceptibility gene: cloning, identification, and sequence.	
#cross-references	MUID:87149066	
#accession	A03152	
#molecule_type	mRNA	
#residues	113-116, 'LLSYRKY', 125-332, 'R', 334-367, 'I', 369-928	
#cross-references	GB:M15400; NID:g190958; PIDN:AAA69807.1; PID:g190959	
REFERENCE	A91613	
#authors	McGee, T.L.; Vandell, D.W.; Dryja, T.P.	
#journal	Gene (1989) 80:119-128	
#title	Structure and partial genomic sequence of the human retinoblastoma susceptibility gene.	
#cross-references	MUID:90006771	
#accession	A91613	
#molecule_type	DNA	
#residues	1-928	
#cross-references	GB:M27845; GB:L11910; NID:g292420; PIDN:AAA53483.1;	
#note	the authors translated the codon GAA for residue 559 as Gly	


```

QY 300 AGAEGSGPEEP 311
      #domain acyl carrier protein homology #label ACP
      #length 88 #molecular-weight 9428 #checksum 8318

RESULT 3
ENTRY 3
TITLE surface protein rhopty ROP1 precursor - Toxoplasma gondii
ORGANISM #formal_name Toxoplasma gondii
DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
ACCESSIONS A58938; A45644; S37697
REFERENCE A58938
#authors Boothroyd, J.C.
#submission Submitted to GenBank, July 1995
#accession A58938
##status preliminary
##molecule_type mRNA
##residues 1-396 ##label BOO
##cross-references GB:M71274; NID:g897822; PIDN:AAA69859.1; PID:g897823
#note revision to sequence reported in A45644
REFERENCE A45644
#authors Ossorio, P.N.; Schwartzman, J.D.; Boothroyd, J.C.
#journal Mol. Biochem. Parasitol. (1992) 50:1-15
#title A Toxoplasma gondii rhopty protein associated with host cell
#penetration has unusual charge asymmetry.
#cross-references MUID:92178277
#accession A45644
##molecule_type mRNA
##residues 'MACROLLSVQNLFFFLRDIYCTDFT', 1-352, 'FPQR', 358-364, 'R',
366, 'I', 393, 'SP', 396.
#note 'AASLMARRAGPYWAKESRMMDRNTGSMLLDSAKTIVSRKSGVLRS'
##cross-references EMBL:M71274; NID:g897822
#note sequence extracted from NCBI backbone (NCBIN:85178,
NCBIP:85179)
CLASSIFICATION #superfamily surface protein rhopty
KEYWORDS surface antigen
FEATURE 1-21
22-396
#domain signal sequence #status predicted #label SIG\
#product surface protein rhopty #status predicted
#label MAT
#length 396 #molecular-weight 42670 #checksum 2217
Query Match 62.5%; Score 55; DB 1; Length 396;
Best Local Similarity 70.0%; Pred. No. 1.47e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 219 EQEGSGPEPL 228
      |||:::|||||
QY 303 EQEGSGPEEP 312

RESULT 4
ENTRY 4
TITLE acyl carrier protein - Streptomyces cinnamonensis
ORGANISM #formal_name Streptomyces cinnamonensis
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
ACCESSIONS S25078; S18169
REFERENCE S25078
#authors Arrowsmith, T.J.; Malpartida, F.; Sherman, D.H.; Birch, A.;
Hopwood, D.A.; Robinson, J.A.
#journal Mol. Gen. Genet. (1992) 234:254-264
#title Characterisation of acti-homologous DNA encoding polyketide
synthase genes from the monensin producer Streptomyces
cinnamonensis.
#cross-references MUID:92374994
#accession S25078
##molecule_type DNA
##residues 1-88 ##label ARR
##cross-references EMBL:211511; NID:g46799; PIDN:CAA77598.1; PID:g46802
CLASSIFICATION #superfamily acyl carrier protein; acyl carrier protein
homology
FEATURE

QY 9-78
SUMMARY #domain acyl carrier protein homology #label ACP
#length 88 #molecular-weight 9428 #checksum 8318

Query Match 60.2%; Score 53; DB 2; Length 88;
Best Local Similarity 46.2%; Pred. No. 3.67e+00;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 19 AGADESAGPDDIL 31
      |||:::|||||
QY 300 AGAEGSGPEEP 312

RESULT 5
ENTRY 5
TITLE TelB protein - plasmid RK2
ALTERNATE_NAMES klac protein
ORGANISM #formal_name plasmid RK2
DATE 16-Oct-1992 #sequence_revision 16-Oct-1992 #text_change
ACCESSIONS C38178; C42362
REFERENCE C38178
#authors Walter, E.G.; Thomas, C.M.; Ibbotson, J.P.; Taylor, D.E.
#journal J. Bacteriol. (1991) 173:1111-1119
#title Transcriptional analysis, translational analysis, and
sequence of the kila-tellurite resistance region of plasmid
RK2Te(r).
#cross-references MUID:91123183
#accession C38178
##molecule_type DNA
##residues 1-317 #label WAL
##cross-references GB:M62846; GB:M38697; NID:g149152; PID:g149155
#note this sequence is the tellurite resistance-conferring
form of the protein
REFERENCE A42362
#authors Goncharoff, P.; Saadi, S.; Chang, C.H.; Saltman, L.H.;
Figurski, D.H.
#journal J. Bacteriol. (1991) 173:3463-3477
#title Structural, molecular, and genetic analysis of the kila
operon of broad-host-range plasmid RK2.
#cross-references MUID:91258329
#accession C42362
##molecule_type DNA
##residues 1-124, 'SW', 126-317 #label GON
##cross-references GB:M62436; NID:g152517; PID:g152520
#note this sequence is the tellurite sensitivity-conferring
form of the protein
COMMENT The RK2 plasmid is a member of the P-alpha incompatibility group of
plasmids, which replicate and are maintained stably in a wide
variety of gram-negative bacteria.
GENETICS
#genome plasmid
SUMMARY #length 317 #molecular-weight 32396 #checksum 8267

Query Match 60.2%; Score 53; DB 2; Length 317;
Best Local Similarity 45.5%; Pred. No. 3.67e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 153 GGDHDAGPAEP 163
      |||:::|||||
QY 301 GAEQSGPEEP 311

RESULT 6
ENTRY 6
TITLE procyclic acidic repetitive protein A-alpha - Trypanosoma
brucei
ORGANISM #formal_name Trypanosoma brucei
DATE 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change
ACCESSIONS A30229; S21539
REFERENCE A30229
#authors Mowatt, M.R.; Wisdom, G.S.; Clayton, C.E.
#journal Mol. Cell. Biol. (1989) 9:1332-1335

```

```

#title      Variation of tandem repeats in the developmentally regulated
#cross-references MUID:89261740
#accession  A30229
#molecule_type DNA
#residues   1-114 ##label MOW
##cross-references GB:M25787; NID:g341399; PIDN:AAA53283.1; PID:g576546
REFERENCE   S21538
#authors    Vijayarathy, S.; Ernest, I.; Itzhaki, J.; Sherman, D.;
#submission Mowatt, M.R.; Michels, P.A.M.; Clayton, C.E.
#description submitted to the EMBL Data Library, April 1990
#description The genes encoding fructose biphosphate aldolase in
#description trypanosoma brucei are interspersed with related genes.
#accession  S21539
##status   preliminary
##molecule_type DNA
##residues 1-114 ##label VIJ
##cross-references EMBL:X52584; NID:g10480; PIDN:CAA36814.1; PID:g10481
CLASSIFICATION #superfamily procyclic acidic repetitive protein
SUMMARY        #length 114 #molecular-weight 11611 #checksum 2849

Query Match      59.1%; Score 52; DB 2; Length 114;
Best Local Similarity 72.7%; Pred. No. 5.75e-00;
Matches          8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 42 GKERDGPPEP 52
QY 301 GAQESGPPEP 311
      | | | | | | | |
      | | | | | | | |

RESULT 7
ENTRY   D70537 #type complete
TITLE   hypothetical protein Rv1115 - Mycobacterium tuberculosis
        (strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE      17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
D70537
REFERENCE A70500
#authors  Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
          C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry
          III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
          Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
          Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
          Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
          Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
          Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
          Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
          Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal  Nature (1998) 393:537-544
#title    Deciphering the biology of Mycobacterium tuberculosis from
          the complete genome sequence.
#cross-references MUID:98295987
#accession D70537
#status   preliminary; nucleic acid sequence not shown;
          translation not shown
##molecule_type DNA
##residues 1-232 ##label COL
##cross-references GB:295585; GB:ALJ23456; NID:g3261787; PID:e317139;
          PID:g2117209
##experimental_source strain H37Rv
GENETICS
#gene     Rv1115
SUMMARY   #length 232 #molecular-weight 24082 #checksum 2766

Query Match      58.0%; Score 51; DB 2; Length 232;
Best Local Similarity 46.2%; Pred. No. 8.97e-00;
Matches          6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 145 AGEDOPPAPPEP 157
QY 300 AGAEGSGPEEP 312
      | | | | | | | |
      | | | | | | | |

```

```

RESULT 8
ENTRY   A40701 #type complete
TITLE   tenascin-X precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE      10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
          29-Sep-1999
ACCESSIONS A40701; A33725; C42175
REFERENCE   A40701
#authors    Bristow, J.; Tee, M.K.; Gitelman, S.E.; Mellon, S.H.; Miller,
          W.L.
#journal    J. Cell Biol. (1993) 122:265-278
#title      tenascin-X: a novel extracellular matrix protein encoded by
          the human XB gene overlapping P450c21B.
#cross-references MUID:93300909
#accession  A40701
##status   preliminary
##molecule_type DNA
##residues 1-3566 ##label BRI
##cross-references EMBL:X71937
REFERENCE   A33725
#authors    Morel, Y.; Bristow, J.; Gitelman, S.E.; Miller, W.L.
#journal    Proc. Natl. Acad. Sci. U.S.A. (1989) 86:6582-6586
#title      Transcript encoded on the opposite strand of the human
          steroid 21-hydroxylase/complement component C4 gene locus.
#cross-references MUID:89367293
#accession  A33725
##molecule_type mRNA
##residues 2748-3199, 'V', 3201-3298, 'E', 3299-3314, 'G', 3316-3566
##cross-references GB:M25813; NID:g183069; PIDN:AAA35884.1; PID:g183070
REFERENCE   A42175
#authors    Matsumoto, K.; Arai, M.; Ishihara, N.; Ando, A.; Inoko, H.;
          Ikemura, T.
#journal    Genomics (1992) 12:485-491
#title      Cluster of fibronectin type III repeats found in the human
          major histocompatibility complex class III region shows the
          highest homology with the repeats in an extracellular
          matrix protein, tenascin.
#cross-references MUID:92217969
#accession  C42175
##molecule_type DNA
##residues 1849-1936 ##label MAT
##experimental_source clone 3.9kF3-1
##note      sequence extracted from NCBI backbone (NCBIP:95694)
GENETICS
#gene      GDB:TNXA; D6S103E; TNX; XA; XB
#map_position 6p21.3-6p21.3
#cross-references GDB:568487; OMIM:600261
CLASSIFICATION #superfamily tenascin-X; EGF homology; fibrinogen beta/gamma
          homology; fibronectin type III repeat homology
          extracellular matrix; glycoprotein; heptad repeat
KEYWORDS
FEATURE     #domain EGF homology #label EGF\
          435-461 #domain fibronectin type III repeat homology #label 3F1\
          748-828 #domain fibronectin type III repeat homology #status
          829-856 atypical #label 3F2\
          873-953 #domain fibronectin type III repeat homology #label 3F3\
          975-1055 #domain fibronectin type III repeat homology #label 3F4\
          1078-1158 #domain fibronectin type III repeat homology #label 3F5\
          1167-1247 #domain fibronectin type III repeat homology #label 3F6\
          1248-1317 #domain fibronectin type III repeat homology #status
          atypical #label 3F7\
          1323-1403 #domain fibronectin type III repeat homology #label 3F8\
          1412-1492 #domain fibronectin type III repeat homology #label 3F9\
          1510-1590 #domain fibronectin type III repeat homology #label
          3F10\
          1618-1676 #domain fibronectin type III repeat homology #status
          atypical #label 3F11\
          1678-1749 #domain fibronectin type III repeat homology #status
          atypical #label 3F12\
          1751-1831 #domain fibronectin type III repeat homology #label
          3F13\

```



```

##molecule_type mrna
##residues 1-316 ##label YAN
##cross-references EMBL:X61171; NID:g312892; PIDN:CAA43479.1;
PID:g312893
CLASSIFICATION #superfamily apolipoprotein A-I
SUMMARY #length 316 #molecular-weight 35870 #checksum 6340

Query Match 55.7%; Score 49; DB 2; Length 316;
Best Local Similarity 69.2%; Pred. No. 2.15e+01;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 18 ADMEGELGPEPL 30
QY 300 AGAEQESGPEPL 312

RESULT 12
ENTRY #type complete
TITLE apolipoprotein E - bovine
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
13-Aug-1999
ACCESSIONS S26478
REFERENCE #authors Brzozowska, A.; Fries, R.; Womack, J.; Grimholt, U.;
Myklebost, O.; Rogne, S.
#submission submitted to the EMBL Data Library, September 1992
#description Isolation, sequencing and expression analysis of a bovine
apolipoprotein E (APOE) cDNA and chromosomal localization
of the APOE locus.
#accession S26478
##status preliminary
##molecule_type mrna
##residues 1-316 ##label BRZ
##cross-references EMBL:X64839; NID:g80; PIDN:CAA46051.1; PID:g81
CLASSIFICATION #superfamily apolipoprotein A-I
SUMMARY #length 316 #molecular-weight 35980 #checksum 6677

Query Match 55.7%; Score 49; DB 2; Length 316;
Best Local Similarity 69.2%; Pred. No. 2.15e+01;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 18 ADMEGELGPEPL 30
QY 300 AGAEQESGPEPL 312

RESULT 13
ENTRY #type complete
TITLE Probable ffh protein - Mycobacterium tuberculosis (strain
H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
23-Jul-1999
ACCESSIONS D70747
REFERENCE #authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession D70747
##status preliminary; nucleic acid sequence not shown;
translation not shown

```

```

##molecule_type DNA
##residues 1-525 ##label COL
##cross-references GB:Z74697; GB:AL123456; NID:g3261602;
PIDN:CAA98978.1; PID:e249960; PID:g1405951
GENETICS #experimental_source strain H37Rv
#gene ffh
CLASSIFICATION #superfamily signal recognition particle 54K protein
SUMMARY #length 525 #molecular-weight 55001 #checksum 888

Query Match 55.7%; Score 49; DB 2; Length 525;
Best Local Similarity 53.8%; Pred. No. 2.15e+01;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 163 PGASPEGGPGDPV 175
QY 300 AGAEQESGPEPL 312

RESULT 14
ENTRY #type complete
TITLE Band 3 anion transport protein, erythrocyte - human
ALTERNATE_NAMES carrier family 4, anion exchanger, member 1; erythroid anion
exchange protein
ORGANISM #formal_name Homo sapiens #common_name man
DATE 03-Aug-1984 #sequence_revision 03-Oct-1995 #text_change
22-Jun-1999
ACCESSIONS A36218; S03074; I39408; A92237; A26507; A92430;
A90323; A28079; S05523; A35835; A44116; A49717;
PC4403; A03189
REFERENCE #accession A36218
#authors Lux, S.E.; John, K.M.; Kopito, R.R.; Lodish, H.F.
#journal Proc. Natl. Acad. Sci. U.S.A. (1989) 86:9089-9093
#title Cloning and characterization of band 3, the human erythrocyte
anion-exchange protein (AE1).
#cross-references MUID:90083213
#accession A36218
##status preliminary
##molecule_type mrna
##residues 1-911 ##label LUX
##cross-references GB:M27819; NID:g178215; PIDN:AAA35514.1; PID:g178216
S03074
REFERENCE #authors Tanner, M.J.A.; Martin, P.G.; High, S.
#journal Biochem. J. (1988) 256:703-712
#title The complete amino acid sequence of the human erythrocyte
membrane anion-transport protein deduced from the cDNA
sequence.
#cross-references MUID:89134172
#accession S03074
##molecule_type mrna
##residues 1-55,'E','57-911 ##label TAN
##cross-references EMBL:X12609; NID:g28713; PIDN:CAA31128.1; PID:g28714
I39408
REFERENCE #authors Showe, L.C.; Ballantine, M.; Huebner, K.
#journal Genomics (1987) 1:71-76
#title Localization of the gene for the erythroid anion exchange
protein, band 3 (EMPB3), to human chromosome 17.
#cross-references MUID:88031311
#accession I39408
##molecule_type DNA
##residues 37-56 ##label SHO1
##cross-references GB:M16978; NID:g178217; PIDN:AAA51670.1; PID:g178220
I39409
#accession I39409
##molecule_type DNA
##residues 118-161 ##label SHO2
##cross-references GB:M16979; NID:g178218; PIDN:AAA51671.1; PID:g553169
A92237
REFERENCE #authors Drickamer, L.K.
#journal J. Biol. Chem. (1978) 253:7242-7248
#title Orientation of the band 3 polypeptide from human erythrocyte
membranes. Identification of NH-2-terminal sequence and
site of carbohydrate attachment.
#cross-references MUID:79027186

```


#accession A92237
#molecule_type protein
#residues 1-3 #label DRI
REFERENCE
#authors Mawby, W.J.; Findlay, J.B.C.
#journal Biochem. J. (1982) 205:465-475
#title Characterization and partial sequence of di-iodosulphophenyl isothiocyanate-binding peptide from human erythrocyte anion-transport protein.
#cross-references MUID:83074521
#accession A26507
#molecule_type protein
#residues 437-473:360-364,'D',366-369 #label MAW
REFERENCE
#authors Kaul, R.K.; Murthy, S.N.P.; Reddy, A.G.; Steck, T.L.; Kohler, H.
#journal J. Biol. Chem. (1983) 258:7981-7990
#title Amino acid sequence of the N(alpha)-terminal 201 residues of human erythrocyte membrane band 3.
#cross-references MUID:83238395
#accession A92430
#molecule_type protein
#residues 1-10,'D',12-68,'E',69-200 #label KAU
REFERENCE
#authors Brock, C.J.; Tanner, M.J.A.; Kempf, C.
#journal Biochem. J. (1983) 213:577-586
#title The human erythrocyte anion-transport protein.
#cross-references MUID:83308584
#accession A90323
#molecule_type protein
#residues 559-630 #label BRO
#note Lys-590 was shown to bind phenyl isothiocyanate, an inhibitor of anion transport
REFERENCE
#authors Kawano, Y.; Okubo, K.; Tokunaga, F.; Miyata, T.; Iwanaga, S.; Hamasaki, N.
#journal J. Biol. Chem. (1988) 263:8232-8238
#title Localization of the pyridoxal phosphate binding site at the COOH-terminal region of erythrocyte band 3 protein.
#cross-references MUID:88228050
#accession A28079
#molecule_type protein
#residues 834-842,'X',844-911 #label KAW
#note Lys-851 was shown to bind the affinity label pyridoxal phosphate, a substrate for anion transport
REFERENCE
#authors Yannoukakos, D.; Vasseur, C.; Blouquit, Y.; Bursaux, E.; Wajcman, H.
#journal Biochim. Biophys. Acta (1989) 998:43-49
#title Primary structure of the cytoplasmic domain of human erythrocyte protein band 3. Comparison with its sequence in the mouse.
#cross-references MUID:90001294
#accession S05523
#molecule_type protein
#residues 1-201:220-292:307-308,'R',310-312,'S',314-329,'K',331-333:347-370 #label YAN1
REFERENCE
#authors Cobb, C.E.; Beth, A.H.
#journal Biochemistry (1990) 29:8283-8290
#title Identification of the eosinyl-5-maleimide reaction site on the human erythrocyte anion-exchange protein: overlap with the reaction sites of other chemical probes.
#cross-references MUID:91070049
#accession A35835
#status Preliminary
#molecule_type protein
#residues 361-364,'X',366-372:424-429,'X',431-434 #label COB
#note Lys-430 is labeled by eosinyl-5-maleimide (EMA) in intact erythrocytes
REFERENCE
#authors Yannoukakos, D.; Vasseur, C.; Driancourt, C.; Blouquit, Y.; Delaunay, J.; Wajcman, H.; Bursaux, E.

Blood (1991) 78:1117-1120
Human erythrocyte band 3 polymorphism (band 3 Memphis): characterization of the structural modification (Lys 56->Glu) by protein chemistry methods.
#cross-references MUID:91329825
#accession A49933
#molecule_type protein
#residues 1-55,'E',57-69 #label YAN2
#note sequence extracted from NCBI backbone (NCBIP:49829)
#note sequence of a common polymorphic form designated band 3 Memphis
A44116
Kang, D.; Okubo, K.; Hamasaki, N.; Kuroda, N.; Shiraki, H.
J. Biol. Chem. (1992) 267:19211-19217
#journal A structural study of the membrane domain of band 3 by tryptic digestion. Conformational change of band 3 in situ induced by alkali treatment.
#cross-references MUID:92406862
#accession A44116
#status Preliminary
#molecule_type protein
#residues 361-372:390-399:604-613:632-639:647-656:699-729:731-743:761-781:818-826 #label KAN
#experimental_source erythrocyte
#note sequence extracted from NCBI backbone (NCBIP:113819, NCBIP:113821, NCBIP:113823, NCBIP:113825, NCBIP:113827, NCBIP:113829, NCBIP:113831, NCBIP:113833, NCBIP:113835)
#note a histidine residue essential for anion transport is suggested to be His-651, His-734, or His-819
A49717
Okubo, K.; Kang, D.; Hamasaki, N.; Jennings, M.L.
J. Biol. Chem. (1994) 269:1918-1926
#journal Red blood cell band 3. Lysine 539 and lysine 851 react with the same H-2DIDS (4,4'-diisothiocyanodihydrostilbene-2,2'-disulfonic acid) molecule.
#cross-references MUID:94124538
#accession A49717
#molecule_type protein
#residues 427-436:479-519:538-540:559-566:809-817:825-841:849-857 #label OKU
#note Lys-539 and Lys-851 were shown to bind the same molecule of the anion transport inhibitor 4, 4'-diisothiocyanostilbene-2,2'-disulfonic acid (H-2DIDS)
PC4403
Hamasaki, N.; Okubo, K.; Kuma, H.; Kang, D.; Yae, Y.
J. Biochem. (1997) 122:577-585
#journal Proteolytic cleavage sites of band 3 protein in alkali-treated membranes: Fidelity of hydropathy prediction for band 3 protein.
#cross-references MUID:98006310
#accession PC4403
#molecule_type protein
#residues 361-911 #label HAM
COMMENT Band 3 is the major integral glycoprotein of the erythrocyte membrane. A dimer in solution, it spans the membrane asymmetrically and appears to be tetrameric.
COMMENT Band 3 has at least two functional domains. Its integral domain mediates a 1:1 exchange of chloride and bicarbonate anions across the membrane, whereas its cytoplasmic domain provides binding sites for cytoskeletal proteins, glycolytic enzymes, and hemoglobin.
GENETICS
#gene GDB:SLC4A1; EPB3
#cross-references GDB:119874; OMIM:109270
#map_position 17q21-17q22
CLASSIFICATION #superfamily band 3 anion transport protein
KEYWORDS acetylated amino end; glycoprotein; ion transport; phosphoprotein; transmembrane protein
FEATURE
1-403 #region cytoskeletal protein binding
404-911 #region anion antiporter

```

405-427      #domain transmembrane #status predicted #label TM01\
...
Note: remainder of annotations omitted.

Query Match      55.7%; Score 49; DB 1; Length 911;
Best Local Similarity 41.7%; Pred. NO. 2.15e+01;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 361 GDLNGGPDPL 372
      | : :||: ||
QY 301 GAQESGPEPL 312

RESULT 15
ENTRY      A33718      #type complete
TITLE      retinoblastoma protein - mouse
ORGANISM   #formal_name Mus musculus #common_name house mouse
DATE       09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change
          13-Aug-1999
ACCESSIONS A33718
REFERENCE   A33718
#authors   Bernards, R.; Schackleford, G.M.; Gerber, M.R.; Horowitz,
          J.M.; Friend, S.H.; Schartl, M.; Bogenmann, E.; Rapaport,
          J.M.; Mc Gee, T.; Dryja, T.P.; Weinberg, R.A.
#journal   Proc. Natl. Acad. Sci. U.S.A. (1989) 86:6474-6478
#title     Structure and expression of the murine retinoblastoma gene
          and characterization of its encoded protein.
#cross-references MUID:89367271
#accession A33718
#status     preliminary
#molecule_type mRNA
#residues   1-921 #label BER
#cross-references GB:M26391; NID:g200452; PIDN:AAA39964.1; PID:g200453
CLASSIFICATION #superfamily retinoblastoma-associated protein
KEYWORDS      DNA binding; nucleus; transcription regulation
SUMMARY       #length 921 #molecular-weight 105337 #checksum 5525

Query Match      55.7%; Score 49; DB 2; Length 921;
Best Local Similarity 66.7%; Pred. NO. 2.15e+01;
Matches 8; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 27 PAQDSGPEPL 38
      | : ||||| |
QY 302 AEQESGPEE-PL 312

Search completed: Wed May 10 14:05:20 2000
Job time : 9 secs.

```

W P S R E H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 13:58:54 2000; MasPar time 90.55 Seconds
Tabular output not generated. 4.709 Million cell updates/sec

Title: >US-09-376-430-2
Description: (300-313) from US09376430A.pep (22 of 25)
Perfect Score: 88
Sequence: 1 AGAQESGPPEPLV 14

Scoring table: PAM 150
Gap 11

Searched: 83857 segs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 24.076; Variance 29.392; scale 0.819

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	61	69.3	1 RB_HUMAN	RETINOBLASTOMA-ASSOCIA	3.12e-02
2	53	60.2	1 ACP_STRCM	MONENSIN POLYKETIDE SY	1.72e+00
3	53	60.2	1 KLAC_ECOLI	KLAC PROTEIN (TELB PRO	1.72e+00
4	53	60.2	1 YWO3_CABEL	HYPOTHETICAL 63_6 KD P	1.72e+00
5	52	59.1	1 PARA_TRYBB	PROCYCLIC FORM SPECIFI	2.75e+00
6	50	56.8	1 MYPC_MOUSE	MYOSIN-BINDING PROTEIN	7.01e+00
7	49	55.7	1 APE_BOVIN	APOLIPOPROTEIN E PRECU	1.11e+01
8	49	55.7	1 HFHA_HUMAN	HEPATOCYTE NUCLEAR FAC	1.11e+01
9	49	55.7	1 SR34_MYCTU	SIGNAL RECOGNITION PAR	1.11e+01
10	49	55.7	1 B3AT_HUMAN	BAND 3 ANION TRANSPORT	1.11e+01
11	49	55.7	1 RB_MOUSE	RETINOBLASTOMA-ASSOCIA	1.11e+01
12	49	55.7	1 ANK1_MOUSE	ANKYRIN.	1.11e+01
13	49	55.7	1 POLN_RUBVT	NONSTRUCTURAL POLYPROT	1.11e+01
14	48	54.5	1 UBC4_CANAL	UBIQUITIN-CONJUGATING	1.73e+01
15	48	54.5	1 UBC5_YEAST	UBIQUITIN-CONJUGATING	1.73e+01
16	48	54.5	1 UBC4_YEAST	UBIQUITIN-CONJUGATING	1.73e+01
17	48	54.5	1 DRG1_RAT	HOMEOBOX PROTEIN DRG11	1.73e+01
18	48	54.5	1 TRP1_HUMAN	TROPONIN T, SLOW SKELE	1.73e+01
19	48	54.5	1 SR54_MCLE	SIGNAL RECOGNITION PAR	1.73e+01
20	48	54.5	1 ZYX_HUMAN	ZYXIN (ZYXIN 2).	1.73e+01
21	48	54.5	1 APB3_HUMAN	AMYLOID BETA A4 PRECUR	1.73e+01
22	48	54.5	1 PUR2_YARLI	BIFUNCTIONAL PURINE BI	1.73e+01
23	48	54.5	1 FBL2_MOUSE	FIBULIN-2 PRECURSOR.	1.73e+01

24	48	54.5	1873	1 CCAS_HUMAN	VOLTAGE-DEPENDENT L-TY	1.73e+01
25	48	54.5	2325	1 PGG2_RAT	CHONDROITIN SULFATE PR	1.73e+01
26	48	54.5	2647	1 ABP2_HUMAN	ENDOTHELIAL ACTIN-BIND	1.73e+01
27	47	53.4	118	1 UBCB_ARATH	UBIQUITIN-CONJUGATING	2.70e+01
28	47	53.4	121	1 UBCB_ARATH	UBIQUITIN-CONJUGATING	2.70e+01
29	47	53.4	147	1 UBCA_HUMAN	UBIQUITIN-CONJUGATING	2.70e+01
30	47	53.4	147	1 UBC1_DROME	UBIQUITIN-CONJUGATING	2.70e+01
31	47	53.4	147	1 UBC1_DROME	UBIQUITIN-CONJUGATING	2.70e+01
32	47	53.4	148	1 UBC8_HUMAN	UBIQUITIN-CONJUGATING	2.70e+01
33	47	53.4	148	1 UBC8_ARATH	UBIQUITIN-CONJUGATING	2.70e+01
34	47	53.4	148	1 UBC9_ARATH	UBIQUITIN-CONJUGATING	2.70e+01
35	47	53.4	148	1 UBC4_ARATH	UBIQUITIN-CONJUGATING	2.70e+01
36	47	53.4	230	1 NIFO_RHISN	UBIQUITIN-CONJUGATING	2.70e+01
37	47	53.4	304	1 RS3_HALMA	NIFQ PROTEIN HOMOLOG.	2.70e+01
38	47	53.4	333	1 PIX2_CHICK	30S RIBOSOMAL PROTEIN	2.70e+01
39	47	53.4	396	1 PSD4_DROME	PITUITARY HOMEOBOX 2 (2.70e+01
40	47	53.4	617	1 NTNO_HUMAN	26S PROTEASOME REGULAT	2.70e+01
41	47	53.4	652	1 RPSD_CAUCR	SODIUM-DEPENDENT NORAD	2.70e+01
42	47	53.4	678	1 PTR3_YEAST	RNA POLYMERASE SIGMA F	2.70e+01
43	47	53.4	726	1 PSC2_ARATH	PIR3 PROTEIN (SSY3 PRO	2.70e+01
44	47	53.4	739	1 UL47_HSVBC	DELTA 1-PYRROLINE-5-CA	2.70e+01
45	47	53.4	899	1 RB_RAT	80.7 KD ALPHA TRANS-IN	2.70e+01
					RETINOBLASTOMA-ASSOCIA	2.70e+01

ALIGNMENTS

RESULT	ID	RB_HUMAN	STANDARD	PRT	928 AA.
AC	P06400	P78499			
DT	01-JAN-1988	(Rel. 06, Created)			
DT	01-JAN-1990	(Rel. 13, Last sequence update)			
DT	15-JUL-1999	(Rel. 38, Last annotation update)			
DE	RETINOBLASTOMA-ASSOCIATED PROTEIN (PP110) (P105-RB) (RB).				
GN	RBI.				

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88014238.

RA Lee W.-H., Shew J.-Y., Hong F.D., Sery T.W., Donoso L.A., Young L.-J.,
Bookstein R., Lee E.Y.-H.P.;
RT "The retinoblastoma susceptibility gene encodes a nuclear
phosphoprotein associated with DNA binding activity."
RL Nature 329:642-645(1987).
RN [2]
RP REVISIONS.

RX MEDLINE; 87149066.
RA Lee W.-H., Bookstein R., Hong F.D., Young L.-J., Shew J.-Y.,
Lee E.Y.-H.P.;
RT "Human retinoblastoma susceptibility gene: cloning, identification,
and sequence."
RL Science 235:1394-1399(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88097427.
RA Friend S.H., Horowitz J.M., Gerber M.R., Wang X.-F., Bogenmann E.,
Li F.P., Weinberg R.A.;
RT "Deletions of a DNA sequence in retinoblastomas and mesenchymal
tumors: organization of the sequence and its encoded protein."
RL Proc. Natl. Acad. Sci. U.S.A. 84:9059-9063(1987).
RN [4]

RX MEDLINE; 90006771.
RA McGee T.L., Yandell D.W., Dryja T.P.;
RT "Structure and partial genomic sequence of the human retinoblastoma
susceptibility gene."
RL Gene 80:119-128(1989).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94063891.
RA Toguchida J., McGee T.L., Ciccarelli J.A., Eagle J.R., Yandell D.W.,
Dryja T.P.;

RT "Complete genomic sequence of the human retinoblastoma susceptibility
RL gene."; *Genomics* 17:535-543(1993).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE-CARCINOMA;
RX MEDLINE; 92319557.
RA Hogg A., Onadim Z., Baird P.N., Cowell J.K.;
RT "Detection of heterozygous mutations in the Rb1 gene in
RT retinoblastoma patients using single-strand conformation polymorphism
RT analysis and polymerase chain reaction sequencing."; *Oncogene* 7:1445-1451(1992).
RL [7]
RN [7]
RP SEQUENCE OF 1-45 FROM N.A.
RX MEDLINE; 89239464.
RA T'Ang A., Wu K.J., Hashimoto T., Liu W.Y., Takahashi R., Shi X.H.,
RA Mihara K., Zhang F.H., Chen Y.Y., Du C., Qian J., Lin Y.G.,
RA Murphree A.L., Qiu W.R., Thompson T., Benedict W.F., Fung Y.K.T.;
RT "Genomic organization of the human retinoblastoma gene."; *Oncogene* 4:401-407(1989).
RL [8]
RN [8]
RP COMPLEX WITH SV40 LARGE T ANTIGEN.
RX MEDLINE; 88270506.
RA Decaprio J.A., Ludlow J.W., Figge J., Shew J.-Y., Huang C.-M.,
RA Lee W.-H., Marsilio E., Paucha E., Livingston D.M.;
RT "SV40 large tumor antigen forms a specific complex with the product of
RT the retinoblastoma susceptibility gene."; *Cell* 54:275-283(1988).
RL [9]
RN [9]
RP PHOSPHORYLATION SITES.
RX MEDLINE; 92097548.
RA Lees J.A., Buckkovich K.J., Marshak D.R., Anderson C.W., Harlow E.;
RT "The retinoblastoma protein is phosphorylated on multiple sites by
RT human cdc2."; *EMBO J.* 10:4279-4290(1991).
RL [10]
RN [10]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 378-562.
RX MEDLINE; 97290453.
RA Kim H.-Y., Cho Y.;
RT "Structural similarity between the pocket region of retinoblastoma
RT tumour suppressor and the cyclin-box."; *Nat. Struct. Biol.* 4:390-395(1997).
RL [11]
RN [11]
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 380-785.
RX MEDLINE; 98154728.
RA Lee J.O., Russo A.A., Pavletich N.P.;
RT "Structure of the retinoblastoma tumour-suppressor pocket domain bound
RT to a peptide from HPV E7."; *Nature* 391:859-865(1998).
RL [12]
RN [12]
RP VARIANT RB LEU-567.
RX MEDLINE; 90081757.
RA Randall D.W., Campbell T.A., Dayton S.H., Petersen R., Walton D.,
RA Little J.B., McConkie-Rosell A., Buckley E., Dryja I.;
RT "Oncogenic point mutations in the human retinoblastoma gene: their
RT application to genetic counseling."; *New Engl. J. Med.* 321:1689-1695(1989).
RL [13]
RN [13]
RP VARIANT RB TRP-661.
RX MEDLINE; 92335261.
RA Onadim Z., Hogg A., Baird P.N., Cowell J.K.;
RT "Oncogenic point mutations in exon 20 of the Rb1 gene in families
RT showing incomplete penetrance and mild expression of the
RT retinoblastoma phenotype."; *Proc. Natl. Acad. Sci. U.S.A.* 89:6177-6181(1992).
RL [14]
RN [14]
RP VARIANT RB ARG-457.
RX MEDLINE; 93348271.
RA Hogg A., Bia B., Onadim Z., Cowell J.K.;
RT "Molecular mechanisms of oncogenic mutations in tumors from patients
RT with bilateral and unilateral retinoblastoma."; *Proc. Natl. Acad. Sci. U.S.A.* 90:7351-7355(1993).
RL [15]
RN [15]

RP VARIANTS RB ASN-480 DEL AND TRP-661.
RX MEDLINE; 95012220.
RA Lohmann D.R., Brandt B., Hoepfing W., Passarge E., Horsthemke B.;
RT "Distinct Rb1 gene mutations with low penetrance in hereditary
RT retinoblastoma."; *Hum. Genet.* 94:349-354(1994).
RL [16]
RN [16]
RP VARIANTS RB GLN-72; TYR-549 AND LYS-803.
RX MEDLINE; 96187126.
RA Liu Z., Song Y., Bia B., Cowell J.K.;
RT "Germline mutations in the Rb1 gene in patients with hereditary
RT retinoblastoma."; *Genes Chromosomes Cancer* 14:277-284(1995).
RL [17]
RN [17]
RP VARIANTS RB THR-185; PRO-635; VAL-654 AND PRO-685.
RX MEDLINE; 95315934.
RA Blanquet V., Turleau C., Gross-Morand M.S., Senamaud-Beaufort C.,
RA Doz F., Besmond C.;
RT "Spectrum of germline mutations in the Rb1 gene: a study of 232
RT patients with hereditary and non hereditary retinoblastoma."; *Hum. Mol. Genet.* 4:383-388(1995).
RL [18]
RN [18]
RP VARIANTS RB ASP-137 AND TRP-661.
RX MEDLINE; 97456418.
RA Lohmann D.R., Gerick M., Brandt B., Oelschlaeger U., Lorenz B.,
RA Passarge E., Horsthemke B.;
RT "Constitutional Rb1-gene mutations in patients with isolated
RT unilateral retinoblastoma."; *Am. J. Hum. Genet.* 61:282-294(1997).
RL [19]
RN [19]
RP VARIANTS RB LEU-567; ARG-712 AND PRO-662.
RX Yilmaz S., Horsthemke B., Lohmann D.R.;
RT "Twelve novel Rb1 gene mutations in patients with hereditary
RT retinoblastoma."; *Hum. Mutat.* 12:434-434(1998).
RL [20]
RN [20]
RP FUNCTION: PROBABLY ACTS AS A REGULATOR OF OTHER GENES. FORMS A
CC COMPLEX WITH ADENOVIRUS E1A AND WITH SV40 LARGE T ANTIGEN. ACTS AS
CC A TUMOR SUPPRESSOR. MAY BIND AND MODULATE FUNCTIONALLY CERTAIN
CC CELLULAR PROTEINS WITH WHICH T AND E1A COMPETE FOR POCKET BINDING.
CC POTENT INHIBITOR OF E2F-MEDIATED TRANS-ACTIVATION, INTERACTS
CC PREFERENTIALLY WITH TRANSCRIPTION FACTOR E2F1.
CC -! SUBCELLULAR LOCATION: NUCLEAR.
CC -! PTM: PHOSPHORYLATED FROM S TO M PHASE OF THE CELL CYCLE AND IS
CC UNPHOSPHORYLATED IN G1. T, BUT NOT E1A, BINDS ONLY TO THE
CC UNPHOSPHORYLATED FORM.
CC -! DISEASE: DELETION OR ALTERATION OF THIS PROTEIN RESULTS IN THE
CC CHILDHOOD CANCER RETINOBLASTOMA (RB). RB IS A CONGENITAL
CC MALIGNANT TUMOR THAT ARISES FROM THE NUCLEAR LAYERS OF THE
CC RETINA. IT OCCURS IN ABOUT 1:20,000 LIVE BIRTHS AND REPRESENTS
CC ABOUT 2% OF CHILDHOOD MALIGNANCIES. IT IS BILATERAL IN ABOUT 30%
CC OF CASES. ALTHOUGH MOST RB APPEAR SPORADICALLY, ABOUT 20% ARE
CC TRANSMITTED AS AN AUTOSOMAL DOMINANT TRAIT WITH INCOMPLETE
CC PENETRANCE. THE DIAGNOSIS IS USUALLY MADE BEFORE THE AGE OF 2
CC YEARS WHEN STRABISMUS OR A GREY TO YELLOW REFLEX FROM PUPIL ("CAT
CC EYE") IS INVESTIGATED.
CC -! SIMILARITY: BELONGS TO THE RETINOBLASTOMA PROTEIN (RB) FAMILY.
CC -! DATABASE: NAME=RB1base; NOTE=RB1 mutation db;
CC WWW="http://home.kamp.net/home/dr.lohmann/".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; L41870; AAB59465.1;
DR EMBL; M27866; AAA53484.1;
DR EMBL; M27845; AAA53484.1; JOINED.
DR EMBL; M27846; AAA53484.1; JOINED.
DR EMBL; M27847; AAA53484.1; JOINED.
DR EMBL; M27849; AAA53484.1; JOINED.

DR EMBL: M27850; AAA53484.1; JOINED.
DR EMBL: M27851; AAA53484.1; JOINED.
DR EMBL: L35146; AAA53484.1; JOINED.
DR EMBL: M27852; AAA53484.1; JOINED.
DR EMBL: M27853; AAA53484.1; JOINED.
DR EMBL: M27854; AAA53484.1; JOINED.
DR EMBL: M27855; AAA53484.1; JOINED.
DR EMBL: M27856; AAA53484.1; JOINED.
...

Note: remainder of annotations omitted.

Query Match 69.3%; Score 61; DB 1; Length 928;
Best Local Similarity 69.2%; Pred. No. 3.12e-02;
Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db 33 PEODSGPDLPLV 45
QY 302 AEQESGPEE-PLV 313
:|||||:|

RESULT 2
ID ACP_STROM STANDARD; PRT; 88 AA.
AC P41174;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE MONENSIN POLYKETIDE SYNTHASE ACYL CARRIER PROTEIN.
OS Streptomyces cinnamonensis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3823.5;
RX MEDLINE; 92374994.
RA Arrowsmith T.J., Malpartida F., Sherman D.H., Birch A., Hopwood D.A.,
RA Robinson J.A.;
RT "Characterisation of actI-homologous DNA encoding polyketide synthase
genes from the monensin producer Streptomyces cinnamonensis.";
RL Mol. Gen. Genet. 234:254-264(1992).
CC -!- FUNCTION: ACYL CARRIER PROTEIN.
CC -!- PATHWAY: BIOSYNTHESIS OF THE POLYKETIDE ANTIBIOTIC MONENSIN.
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: Z11511; CAA77598.1; -
PIR: S25078; S25078.
DR HSP; Q02054; IAF8.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
DR PROSITE; PS00075; ACP_DOMAIN; 1.
DR PFAM; PF00550; pp-binding; 1.
KW Antibiotic biosynthesis; Phosphopantetheine.
FT BINDING 42 42
FT PHOSPHOPANTHETHEINE (POTENTIAL).
QY SEQUENCE 88 AA; 9428 MW; A62AA9FC529C71E8 CRC64;

Query Match 60.2%; Score 53; DB 1; Length 88;
Best Local Similarity 46.2%; Pred. No. 1.72e-00;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 19 AGAESAGPDDL 31
QY 300 AGAESGPEEP 312
:|||||:|

RESULT 3
ID KLAC_ECOLI STANDARD; PRT; 317 AA.
AC Q32329;
DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE KLAC PROTEIN (TELB PROTEIN).
GN KLAC OR TELB.
OS Escherichia coli.
OG Plasmid Incp-alpha RK2.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91258329.
RA Goncharoff P., Saadi S., Chang C.H., Saltman L.H., Figurski D.H.;
RT "Structural, molecular, and genetic analysis of the Kila operon of
broad-host-range plasmid RK2.";
RL J. Bacteriol. 173:3463-3477(1991).
CC -!- FUNCTION: BELONGS TO THE KLA OPERON, WHICH IS ASSOCIATED WITH
CC CRYPTIC TELLURITE RESISTANCE, AND INCW PLASMID FERTILITY
CC INHIBITION.
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: M62426; AAA26403.1; -
PIR: -
KW Plasmid; Tellurium resistance.
QY SEQUENCE 317 AA; 32380 MW; BC6F2C9EF7CCA686 CRC64;

Query Match 60.2%; Score 53; DB 1; Length 317;
Best Local Similarity 45.5%; Pred. No. 1.72e-00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 153 GGDHDGAPAP 163
QY 301 GAQESGPEEP 311
:|||||:|

RESULT 4
ID YW03_CAEEL STANDARD; PRT; 563 AA.
AC Q10906;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 63.6 KD PROTEIN AH9.3 IN CHROMOSOME X.
GN AH9.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Stellyes L., Gattung S.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: U41270; AAA82441.1; -
DR WORMPEP; AH9.3; CE03859.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 221 241
FT TRANSMEM 280 300
FT POTENTIAL.
QY SEQUENCE 563 AA; 63615 MW; 0905EA22C871A8AD CRC64;


```

FT CONFLICT 32 32 SIMILARITY).
FT CONFLICT 39 39 E -> G (IN REF. 2).
FT CONFLICT 113 113 M -> K (IN REF. 2).
FT CONFLICT 249 249 E -> K (IN REF. 2).
FT CONFLICT 291 291 P -> S (IN REF. 2).
FT CONFLICT 339 344 EACHRP -> TDLRGM (IN REF. 2).
FT CONFLICT 659 659 T -> A (IN REF. 2).
FT CONFLICT 691 691 A -> T (IN REF. 2).
FT CONFLICT 738 738 E -> G (IN REF. 2).
FT CONFLICT 923 923 R -> T (IN REF. 2).
SQ SEQUENCE 1270 AA; 140632 MW; 699947C3C9B58931 CRC64;

Query Match 56.8%; Score 50; DB 1; Length 1270;
Best Local Similarity 30.8%; Pred. No. 7.01e+00;
Matches 4; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

Db 140 SAAHQGAPDDPI 152
QY 300 AGAEQSGPEEPL 312

RESULT 7
ID APE_BOVIN STANDARD; PRT; 316 AA.
AC Q03247;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE APOLIPOPROTEIN E PRECURSOR (APO-E).
GN APOE.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 93136592.
RA Brzowska A., Fries R., Womack J.E., Grimholt U., Myklebost O.,
RA Rogne S.;
RT "Isolation, sequencing, and expression analysis of a bovine
RT apolipoprotein E (APOE) cDNA and chromosomal localization of the APOE
RT locus.";
RL Mamm. Genome 4:53-57(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 9132915.
RA Yang I.W., Chan L., Li W.H.;
RT "Cloning and sequencing of bovine apolipoprotein E complementary DNA
RT and molecular evolution of apolipoproteins E, C-I, and C-II.";
RL J. Mol. Evol. 32:469-475(1991).
CC -!- FUNCTION: APO-E MEDIATES BINDING, INTERNALIZATION, AND CATABOLISM
CC OF LIPOPROTEIN PARTICLES. IT CAN SERVE AS A LIGAND FOR THE LDL(APO
CC B/E) RECEPTOR AND FOR THE SPECIFIC APO-E RECEPTOR (CHYLOMICRON
CC REMNANT) OF HEPATIC TISSUES.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- TISSUE SPECIFICITY: SECRETED IN PLASMA.
CC -!- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64839; CAA46051.1; -
DR EMBL; X61171; CAA43479.1; -
DR PIR; S26478; S26478.
DR HSSP; P02649; 10EF.
DR PFWA; PF01442; Apolipoprotein; 1.
DR Glycoprotein; Plasma; Lipid transport; HDL; VLDL; Chylomicron;
KW Heparin-binding; Repeat; Signal.
FT SIGNAL 1 18 BY SIMILARITY.
FT CHAIN 19 316 APOLIPOPROTEIN E.
FT DOMAIN 157 167 LDL RECEPTOR BINDING (BY SIMILARITY).
FT DOMAIN 161 164 HEPARIN-BINDING (BY SIMILARITY).
FT DOMAIN 228 235 HEPARIN-BINDING (BY SIMILARITY).
FT DOMAIN 254 254 8 X 22 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 79 100 1.
FT REPEAT 101 122 2.
FT REPEAT 123 144 3.
FT REPEAT 145 166 4.
FT REPEAT 167 188 5.
FT REPEAT 189 210 6.
FT REPEAT 211 232 7.
FT REPEAT 233 254 8.
FT CONFLICT 129 130 RN -> CG (IN REF. 2).
SQ SEQUENCE 316 AA; 35980 MW; 1231292FC8DA9F1E CRC64;

Query Match 55.7%; Score 49; DB 1; Length 316;
Best Local Similarity 69.2%; Pred. No. 1.11e+01;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 18 ADMEGELGPEEPL 30
QY 300 AGAEQSGPEEPL 312

RESULT 8
ID HF4_HUMAN STANDARD; PRT; 421 AA.
AC Q92949; O00630;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEPATOCTE NUCLEAR FACTOR 3 FORKHEAD HOMOLOG 4 (HFH-4).
GN FKHL13 OR HF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FETAL LUNG;
RX MEDLINE; 98191201.
RA Pelletier G.J., Brody S.L., Liapis H., White R.A., Hackett B.P.;
RT "A human forkhead/winged-helix transcription factor expressed in
RT developing pulmonary and renal epithelium.";
RL Am. J. Physiol. 274:L351-L359(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE; 97230460.
RA Murphy D.B., Seemann S., Wieser S., Kirschner R., Grzeschik K.H.,
RA Thies U.;
RT "The human hepatocyte nuclear factor 3/fork head gene FKHL13: genomic
RT structure and pattern of expression.";
RL Genomics 40:462-469(1997).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN CELL FATE DETERMINATION
CC DURING LUNG DEVELOPMENT AND IN SPERMATOGENESIS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS, OVIDUCT, LUNG, AND BRAIN CORTEX.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN DEVELOPING LUNG, KIDNEY, AND
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U69537; AAB09039.1; -
DR EMBL; X99350; CAA67730.1; -
```

```

DR EMBL; X99351; CAA67730.1; JOINED.
DR EMBL; X99349; CAA67729.1; -.
DR MIM; 602291; -.
DR PRINTS; PR00053; FORKHEAD.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
DR PFAM; PF00250; Fork_head; 1.
KW DNA-binding; Nuclear protein; Transcription regulation.
FT DNA_BIND 120 210 FORK-HEAD.
FT CONFLICT 5 5 W -> C (IN REF. 2).
FT CONFLICT 11 11 A -> T (IN REF. 2).
FT CONFLICT 56 56 G -> A (IN REF. 1).
FT CONFLICT 311 311 W -> L (IN REF. 1).
FT CONFLICT 17 22 AGPEGG -> PAGGR (IN REF. 2).
FT CONFLICT 257 257 A -> P (IN REF. 2).
FT CONFLICT 262 267 GWGAGGRLGHRKQPLPKRVKVP -> VVWQARAGWDI
FT CONFLICT 330 330 V -> L (IN REF. 2).
FT CONFLICT 421 AA; 45232 MW; BC10E03B717C58C5 CRC64;
SQ SEQUENCE 421 AA; 45232 MW; BC10E03B717C58C5 CRC64;

Query Match 55.7%; Score 49; DB 1; Length 421;
Best Local Similarity 53.8%; Pred. No. 1.11e+01;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 11 AGPAEAGPEGGL 23
QY 300 AGAEQESGPEPL 312
II : : : : :

RESULT 9
ID SR54_MYCTU STANDARD; PRT; 525 AA.
AC Q10963;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE SIGNAL RECOGNITION PARTICLE PROTEIN (FIFTY-FOUR HOMOLOG).
GN FFH OR RV2916C OR MTCY338.04C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 98295987.
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES (BY SIMILARITY).
CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A SMALL
CC CYTOPLASMIC RNA (SC-RNA) MOLECULE AND PROTEIN FFH (BY SIMILARITY).
CC -!- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; Z74697; CAA98978.1; -.
DR HSP; O07347; 1FFH.
DR PROSITE; PS00300; SRP54; 1.
DR PFAM; PF00448; SRP54; 1.
DR TUBERCULIST; RV2916C; -.
KW Signal recognition particle; GTP-binding; RNA-binding.
FT DOMAIN 1 301 G-DOMAIN (BY SIMILARITY).
FT NP_BIND 302 525 M-DOMAIN (BY SIMILARITY).
FT NP_BIND 107 114 GTP (BY SIMILARITY).
FT NP_BIND 196 200 GTP (BY SIMILARITY).
FT NP_BIND 254 257 GTP (BY SIMILARITY).
SQ SEQUENCE 525 AA; 55001 MW; 36F6227BC367817C CRC64;

Query Match 55.7%; Score 49; DB 1; Length 525;
Best Local Similarity 53.8%; Pred. No. 1.11e+01;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 163 PGASPESGPDV 175
QY 300 AGAEQESGPEPL 312
II : : : : :

RESULT 10
ID B3AT_HUMAN STANDARD; PRT; 911 AA.
AC P02730;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE BAND 3 ANION TRANSPORT PROTEIN (ANION EXCHANGE PROTEIN 1) (AE 1).
GN SLC4A1 OR AE1 OR EPB3 OR DI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90083213.
RA Lux S.E., John K.M., Kopito R.R., Lodish H.F.;
RT "Cloning and characterization of band 3, the human erythrocyte anion-
RT exchange protein (AE1).";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9089-9093(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BLOOD;
RX MEDLINE; 89134172.
RA Tanner M.J.A., Martin P.G., High S.;
RT "The complete amino acid sequence of the human erythrocyte membrane
RT anion-transport protein deduced from the cDNA sequence.";
RL Biochem. J. 256:703-712(1988).
RN [3]
RP SEQUENCE OF 1-199; 220-292 AND 347-370.
RX MEDLINE; 90001294.
RA Yannoukakis D., Vasseur C., Blouquit Y., Bursaux E., Wajcman H.;
RT "Primary structure of the cytoplasmic domain of human erythrocyte
RT protein band 3. Comparison with its sequence in the mouse.";
RL Biochim. Biophys. Acta 998:43-49(1989).
RN [4]
RP SEQUENCE OF 1-201.
RX MEDLINE; 83238395.
RA Kaul R.K., Murthy S.N.P., Reddy A.G., Steck T.L., Kohler H.;
RT "Amino acid sequence of the N alpha-terminal 201 residues of human
RT erythrocyte membrane band 3.";
RL J. Biol. Chem. 258:7981-7990(1983).
RN [5]
RP SEQUENCE OF 1-3.
RX MEDLINE; 79027186.
RA Drickamer L.K.;
RT "Orientation of the band 3 polypeptide from human erythrocyte
RT membranes. Identification of NH2-terminal sequence and site of
RT carbohydrate attachment.";
RL J. Biol. Chem. 253:7242-7248(1978).
RN [6]
RP SEQUENCE OF 559-630.

```


[15] VARIANT BLOOD GROUP DI(A)/MEMPHIS-II.
RP MEDLINE: 94266802.
RX
RA Bruce L.J., Anstee D.J., Spring F.A., Tanner M.J.;
RT "Band 3 Memphis variant 1f. Altered stibene disulfonate binding and
RT the Diego (Dia) blood group antigen are associated with the human
RT erythrocyte band 3 mutation Pro-854-->Leu.";
RL J. Biol. Chem. 269:16155-16158(1994).
[16]
RN VARIANT BLOOD GROUP WR(A)†
RX MEDLINE: 95111140.
RX Bruce L.J., Ring S.M., Anstee D.J., Reid M.E., Wilkinson S.,
RA Tanner M.J.;
RT "Changes in the blood group Wright antigens are associated with a
RT mutation at amino acid 658 in human erythrocyte band 3: a site of
RT interaction between band 3 and glycophorin A under certain
RT conditions.";
RL Blood 85:541-547(1995).
[17]
RN VARIANTS HS LYS-40; CYS-518 AND MET-663 DEL.
RX MEDLINE: 96225450.
RX Eber S.W., Gonzalez J.M., Lux M.L., Scarpa A.L., Tse W.T.,
RA Dornwell M., Herbers J., Kugler W., Oezcan R., Pekrun A.,
RA Gallagher P.G., Schroeter W., Forget B.G., Lux S.E.;
RT "Ankyrin-1 mutations are a major cause of dominant and recessive
RT hereditary spherocytosis.";
RL Nat. Genet. 13:214-218(1996).
[18]
RN PHOSPHORYLATION SITES.
RP MEDLINE: 91152059.
RX Yannoukakos D., Vasseur C., Piau J.-P., Wajcman H., Bursaux E.;
RT "Phosphorylation sites in human erythrocyte band 3 protein.";
RL Biochim. Biophys. Acta 1061:253-266(1991).
[19]
RN STRUCTURE BY ELECTRON CRYO-MICROSCOPY.
RP MEDLINE: 94320586.
RX Wang D.N., Kuehlbrandt W., Sarabia V.E., Reithmeier R.A.F.;
RT "Two-dimensional structure of the membrane domain of human band 3,
RT the anion transport protein of the erythrocyte membrane.";
RL EMBO J. 12:2233-2239(1993).
[20]
RN STRUCTURE BY ELECTRON CRYO-MICROSCOPY.
RP MEDLINE: 94320586.
RX Wang D.N., Sarabia V.E., Reithmeier R.A.F., Kuehlbrandt W.;
RT "Three-dimensional map of the dimeric membrane domain of the human
RT erythrocyte anion exchanger, Band 3.";
RL EMBO J. 13:3230-3235(1994).
[21]
RN STRUCTURE BY NMR OF 1-16.
RX MEDLINE: 96118374.
RX Schneider M.L., Post C.B.;
RT "Solution structure of a band 3 peptide inhibitor bound to aldolase:
RT a proposed mechanism for regulating binding by tyrosine
RT phosphorylation.";
RL Biochemistry 34:16574-16584(1995).
[22]
RN STRUCTURE BY NMR OF 1-16.
RX MEDLINE: 98118173.
RX Eisenmesser E.Z., Post C.B.;
RT "Insights into tyrosine phosphorylation control of protein-protein
RT association from the NMR structure of a band 3 peptide inhibitor
RT bound to glyceraldehyde-3-phosphate dehydrogenase.";
RL Biochemistry 37:867-877(1998).
[23]
RN STRUCTURE BY NMR OF 389-430.
RX MEDLINE: 98439044.
RX Chambers E.J., Askin D., Bloomberg G.B., Ring S.M., Tanner M.J.;
RT "Studies on the structure of a transmembrane region and a cytoplasmic
RT loop of the human red cell anion exchanger.";
RL Biochem. Soc. Trans. 26:516-520(1998).
[24]
RN STRUCTURE BY NMR OF 405-424 AND 436-456.
RX MEDLINE: 94222092.

RA Gargaro A.R., Bloomberg G.B., Dempsey C.E., Murray M., Tanner M.J.A.;
 RT "The solution structures of the first and second
 RT transmembrane-spanning segments of band 3.";
 RL Eur. J. Biochem. 221:445-454(1994).
 RN [25]

***: remainder of annotations omitted.

Query Match 55.7%; Score 49; DB 1; Length 911;
 Best Local Similarity 41.7%; Pred. No. 1.11e+01;
 Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 361 GLDLNGPDDPL 372
 | : : : : :
 Qy 301 GAQESGPEEL 312

RESULT 11
 ID RB_MOUSE STANDARD; PRT; 921 AA.
 AC P13405;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE RETINOBLASTOMA-ASSOCIATED PROTEIN (P1305) (RB).
 GN RB1 OR RB-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 8936721.
 RA Bernards R., Schackelford G.M., Gerber M.R., Horowitz J.M.,
 RA Friend S.H., Scharf M., Bogenmann E., Rapaport J., McGee T.,
 RA Dryja T., Weinberg R.A.;
 RT "Structure and expression of the murine retinoblastoma gene and
 RT characterization of its encoded protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:6474-6478(1989).
 CC -!- FUNCTION: PROBABLY ACTS AS A REGULATOR OF OTHER GENES. FORMS A
 CC COMPLEX WITH ADENOVIRUS E1A AND WITH SV40 LARGE T ANTIGEN. ACTS AS
 CC A TUMOR SUPPRESSOR. MAY BIND AND MODULATE FUNCTIONALLY CERTAIN
 CC CELLULAR PROTEINS WITH WHICH T AND E1A COMPETE FOR POCKET BINDING.
 CC POTENT INHIBITOR OF E2F-MEDIATED TRANS-ACTIVATION.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- PTM: PHOSPHORYLATED FROM S TO M PHASE OF THE CELL CYCLE AND IS
 CC DEPHOSPHORYLATED IN G1. T, BUT NOT E1A, BINDS ONLY TO THE
 CC UNPHOSPHORYLATED FORM.
 CC -!- SIMILARITY: BELONGS TO THE RETINOBLASTOMA PROTEIN (RB) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M26391; AAA39964.1; -
 CC PIR; A33718; A33718.
 CC HSSP; P06400; 1AD6.
 CC MGD; MGI:97874; RB1.
 CC Transcription regulation; DNA-binding; Nuclear protein;
 KW Phosphorylation; Anti-oncogene.
 FT DOMAIN 9 12 POLY-ALA.
 FT DOMAIN 14 22 POLY-PRO.
 SQ SEQUENCE 921 AA; 105338 MW; 149ED65A07BC6495 CRC64;

Query Match 55.7%; Score 49; DB 1; Length 921;
 Best Local Similarity 66.7%; Pred. No. 1.11e+01;
 Matches 8; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 27: PAQDSGPPEL 38
 | : : : : :
 Qy 302 AEQSGPEE-PL 312

RESULT 12
 ID ANK1_MOUSE STANDARD; PRT; 1862 AA.
 AC Q02357;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE ANKYRIN.
 GN ANK1 OR ANK-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 92345717.
 RA White R.A., Birkenmeier C.S., Peters L.L., Barker J.E., Lux S.E.;
 RT "Murine erythrocyte ankyrin cDNA: highly conserved regions of the
 RT regulatory domain.";
 RL Mamm. Genome 3:281-285(1992).
 CC -!- FUNCTION: ATTACH INTEGRAL MEMBRANE PROTEINS TO CYTOSKELETAL
 CC ELEMENTS; BIND TO THE ERYTHROCYTE MEMBRANE PROTEIN BAND 4.2, TO
 CC NA-K ATPASE, TO THE LYMPHOCYTE MEMBRANE PROTEIN GP85, AND TO THE
 CC CYTOSKELETAL PROTEINS FODRIN, TUBULIN, VIMENTIN AND DESMIN.
 CC ERYTHROCYTE ANKYRINS ALSO LINK SPECTRIN (BETA CHAIN) TO THE
 CC CYTOPLASMIC DOMAIN OF THE ERYTHROCYTES ANION EXCHANGE PROTEIN;
 CC THEY RETAIN MOST OR ALL OF THESE BINDING FUNCTIONS.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF ERYTHROCYTIC
 CC PLASMA MEMBRANE.
 CC -!- PTM: REGULATED BY PHOSPHORYLATION (BY SIMILARITY).
 CC -!- PTM: ACYLATED BY PALMITIC ACID GROUP(S) (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 22 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M84756; AAA37236.1; -
 CC HSSP; Q00420; 1AWC.
 CC MGD; MGI:88024; ANK1.
 CC PROSITE; PS50017; DEATH_DOMAIN; 1.
 CC PFAM; PF00023; ank; 23.
 CC PFAM; PF00531; death; 1.
 CC PFAM; PF00791; ZU5; 1.
 CC Cytoskeleton; Repeat; ANK repeat; Phosphorylation; Lipoprotein.
 FT DOMAIN 1 827 BINDING DOMAIN)
 FT DOMAIN 828 1386 89 KD DOMAIN (ANION EXCHANGE PROTEIN
 FT DOMAIN 1387 1862 62 KD DOMAIN (SPECTRIN BINDING
 FT DOMAIN 55 KD REGULATORY DOMAIN (REGULATES
 FT THE BINDING OF ANKYRIN TO SPECTRIN
 FT AND THE BAND 3 PROTEIN).
 FT 22 X ANK MOTIF REPEATS.
 FT ANK MOTIF 1.
 FT ANK MOTIF 2.
 FT ANK MOTIF 3.
 FT ANK MOTIF 4.
 FT ANK MOTIF 5.
 FT ANK MOTIF 6.
 FT ANK MOTIF 7.
 FT ANK MOTIF 8.
 FT ANK MOTIF 9.
 FT ANK MOTIF 10.
 FT ANK MOTIF 11.
 FT ANK MOTIF 12.
 FT ANK MOTIF 13.
 FT ANK MOTIF 14.
 FT ANK MOTIF 15.

Thu May 11 06:50:15 2000

```
FT REPEAT 531 563 ANK MOTIF 16.
FT REPEAT 564 596 ANK MOTIF 17.
FT REPEAT 597 629 ANK MOTIF 18.
FT REPEAT 630 662 ANK MOTIF 19.
FT REPEAT 663 695 ANK MOTIF 20.
FT REPEAT 696 728 ANK MOTIF 21.
FT REPEAT 729 761 ANK MOTIF 22.
FT DOMAIN 1399 1483 DEATH DOMAIN.
FT SEQUENCE 1862 AA; 204242 MW; AEG85B5B29001E5 CRC64;

Query Match 55.7%; Score 49; DB 1; Length 1862;
Best Local Similarity 42.9%; Pred. No. 1.11e+01;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 1649 SCGEQDTEVSLV 1662
:||||: |::|
QY 300 AGAEGSGPEEPLV 313

RESULT 13
ID POLN_RUBVT STANDARD; PRT: 2205 AA.
AC P13889;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE NONSTRUCTURAL POLYPROTEIN (CONTAINS: NONSTRUCTURAL PROTEINS NSP1 TO NSP4).
OS Rubella virus (strain Therien).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Rubivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90281585.
RA Dominguez G., Wang C.Y., Frey T.K.;
RT "Sequence of the genome RNA of rubella virus: evidence for genetic rearrangement during togavirus evolution.";
RL Virology 177:225-258(1990).
RN [2]
RP SEQUENCE OF 1737-2205 FROM N.A.
RX MEDLINE; 88226020.
RA Frey T.K., Marx L.D.;
RT "Sequence of the region coding for virion proteins C and E2 and the carboxy terminus of the nonstructural proteins of rubella virus: comparison with alphaviruses.";
RL Gene 62:85-99(1988).
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15240; AAA88528.1; -
DR PIR; A35320; MNWVRN.
DR PFAM; PF01661; DUF27; 1.
KW Polyprotein; Nonstructural protein.
FT CHAIN 1 ? NONSTRUCTURAL PROTEIN NSP1.
FT CHAIN 2 ? NONSTRUCTURAL PROTEIN NSP2.
FT CHAIN 3 ? NONSTRUCTURAL PROTEIN NSP3.
FT CHAIN 4 ? NONSTRUCTURAL PROTEIN NSP4.
FT CHAIN 5 ? NONSTRUCTURAL PROTEIN NSP5.
FT SEQUENCE 2205 AA; 240220 MW; 15A95F55E34C0B03 CRC64;

Query Match 55.7%; Score 49; DB 1; Length 2205;
Best Local Similarity 46.2%; Pred. No. 1.11e+01;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 1232 GSEHAPDHLV 1244
:||||: |::|
QY 301 GAEQSGPEEPLV 313
```

```
RESULT 14
ID UBC4_CANAL STANDARD; PRT: 147 AA.
AC P43102;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE UBIQUITIN-CONJUGATING ENZYME E2-16 KD (EC 6.3.2.19) (UBIQUITIN-PROTEIN LIGASE) (UBIQUITIN CARRIER PROTEIN).
DE UBC4.
GN Candida albicans (Yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Candidaceae; Candida.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95212922.
RA Damagnez V., Rolfe M., Cottarel G.;
RT "Schizosaccharomyces pombe and Candida albicans cDNA homologues of the Saccharomyces cerevisiae UBC4 gene.";
RL Gene 155:137-138(1995).
CC -!- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER PROTEINS. MEDIATES THE SELECTIVE DEGRADATION OF SHORT-LIVED AND ABNORMAL PROTEINS.
CC -!- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE -> AMP + PYROPHOSPHATE + PROTEIN N-UBIQUITYLYSINE.
CC -!- PATHWAY: SECOND STEP IN UBIQUITIN CONJUGATION.
CC -!- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-THIOLESTER FORMATION.
CC -!- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L37383; -; NOT_ANNOTATED_CDS.
DR HSP; P15731; 2UC6.
DR PROSITE; PS00183; UBIQUITIN_CONJUGAT; 1.
DR PFAM; PF00179; UQ_con; 1.
KW Ubiquitin conjugation; Ligase; Multigene family.
FT BINDING 85 85 UBIQUITIN (BY SIMILARITY).
FT SEQUENCE 147 AA; 16340 MW; A86357BE37375EB9 CRC64;

Query Match 54.5%; Score 48; DB 1; Length 147;
Best Local Similarity 44.4%; Pred. No. 1.73e+01;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 112 DAMPDDPLV 120
:|:|:|
QY 305 ESGPEEPLV 313

RESULT 15
ID UBC5_YEAST STANDARD; PRT: 148 AA.
AC P15732;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE UBIQUITIN-CONJUGATING ENZYME E2-16 KD (EC 6.3.2.19) (UBIQUITIN-PROTEIN LIGASE) (UBIQUITIN CARRIER PROTEIN).
DE UBC5 OR YDR059C OR YD9609.13C OR D4234.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 40-64 AND 119-125.
RX MEDLINE; 90151631.
RA Seufert W., Jentsch S.;
RT "Ubiquitin-conjugating enzymes UBC4 and UBC5 mediate selective degradation of short-lived and abnormal proteins.";
```

```

RL EMBO J. 9:543-550(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Hunt S., Bowman S., Harris D., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96381250.
RA Brandt P., Ramlow S., Otto B., Bloecker H.;
RT "Nucleotide sequence analysis of a 32,500 bp region of the right arm
RL of Saccharomyces cerevisiae chromosome IV.";
RN yeast 12:85-90(1996).
CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO
CC OTHER PROTEINS. MEDIATES THE SELECTIVE DEGRADATION OF SHORT-LIVED
CC AND ABNORMAL PROTEINS.
CC -1- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE - AMP +
CC PYROPHOSPHATE + PROTEIN N-UBIQUITYLYSINE.
CC -1- PATHWAY: SECOND STEP IN UBIQUITIN CONJUGATION.
CC -1- INDUCTION: ACTIVATED BOTH BY HEAT SHOCK AND CADMIUM.
CC -1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR
CC UBIQUITIN-THIOLESTER FORMATION.
CC -1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17494; CAA35529.1; -
DR EMBL; Z49209; CAA89088.1; -
DR EMBL; X84162; CAA58975.1; -
DR EMBL; Z74355; CAA98877.1; -
DR PIR; S22858; S22858.
DR HSP; P15731; 2UCE.
DR SGD; L0002408; UBC5.
DR PROSITE; PS00183; UBIQUITIN_CONJUGAT; 1.
DR PFAM; PF00179; UQ_con; 1.
KW Ubiquitin conjugation; Ligase; Multigene family.
FT MOD_RES 1 1 BLOCKED.
FT BINDING 86 86 UBIQUITIN (BY SIMILARITY).
SQ SEQUENCE 148 AA; 16280 MW; DC72805EC1F819AE CRC64;

```

```

Query Match          54.5%; Score 48; DB 1; Length 148;
Best Local Similarity 44.4%; Pred. No. 1.73e+01;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 113 DANPDDPLV 121
   :: |::|||
QY 305 ESGPEEPLV 313

```

```

Search completed: Wed May 10 14:00:34 2000
Job time : 100 secs.

```

WATERMAN

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 14:00:52 2000; MasPar time 231.62 Seconds
4.191 Million cell updates/sec

Tabular output not generated.

Title: >US-09-376-430-2
Description: (300-313) from US09376430A.pep (22 of 25)
Perfect score: 88
Sequence: 1 AGAEQESGPPEPLV 14

Scoring table:
PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrmb112

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 23.339; Variance 31.138; scale 0.750

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	61	69.3	53	4	Q92728 RETINOBLASTOMA SUSCEPT	1.38e+01
2	56	63.6	649	3	O94538 HYPOTHETICAL 74.7 KD P	1.50e+00
3	56	63.6	1300	14	O36421 PUTATIVE IMMEDIATE EAR	1.50e+00
4	55	62.5	396	5	O04151 RHOPTRY PROTEIN PRECUR	2.38e+00
5	53	60.2	80	5	P81793 NEUTROXIN TX3-A PRECU	5.92e+00
6	53	60.2	90	5	O18562 IGG AND IGE IMMUNOREAC	5.92e+00
7	53	60.2	128	5	O18563 IGG AND IGE IMMUNOREAC	5.92e+00
8	53	60.2	144	5	O18564 IGG AND IGE IMMUNOREAC	5.92e+00
9	53	60.2	152	5	O18561 IGG AND IGE IMMUNOREAC	5.92e+00
10	53	60.2	156	5	O43394 IGG IMMUNOREACTIVE ANT	5.92e+00
11	53	60.2	317	2	O52327 TELLURITE RESISTANCE.	5.92e+00
12	53	60.2	329	4	O91502 TUMOR SUPPRESSING STF	5.92e+00
13	52	59.1	365	4	O60509 NSP-LIKE 1 (RTN2-B) (F	9.26e+00
14	52	59.1	545	4	O75298 RTN2-A (RTN2-C).	9.26e+00
15	52	59.1	650	10	O941X5 SIMILAR TO EPSTEIN-BAR	9.26e+00
16	51	58.0	232	14	P89459 HYPOTHETICAL 24.1 KD P	1.44e+01
17	51	58.0	3122	14	P89459 VERY LARGE TEGUMENT PR	1.44e+01
18	50	56.8	1183	11	O89040 PHOSPHOLIPASE C BETA 2	2.23e+01
19	50	56.8	3726	11	O61329 AT MOTIF-BINDING FACTO	2.23e+01
20	49	55.7	216	10	O82793 PUTATIVE INORGANIC PYR	3.43e+01

21	49	55.7	243	11	Q61303 ANKYRIN 1, ERYTHROID (3.43e+01
22	49	55.7	283	13	O93417 CASPASE-3.	3.43e+01
23	49	55.7	865	4	O9Y464 TENASCIN-X (FRAGMENT).	3.43e+01
24	49	55.7	988	4	O9Y2E7 KIAA0938 PROTEIN.	3.43e+01
25	49	55.7	1307	5	Q22670 T22C1.10 PROTEIN.	3.43e+01
26	49	55.7	2115	14	O86500 M33 RNA FOR A NONSTRUC	3.43e+01
27	49	55.7	2116	14	O40955 NONSTRUCTURAL PROTEIN.	3.43e+01
28	49	55.7	2588	11	O88491 NSD1 PROTEIN.	3.43e+01
29	49	55.7	4289	4	P78530 TENASCIN X (TENASCIN-X	3.43e+01
30	48	54.5	126	4	Q12966 DIHYDROPYRIDINE RECEPT	5.24e+01
31	48	54.5	147	3	O74196 UBIQUITIN-CONJUGATING	5.24e+01
32	48	54.5	180	11	O35445 HYPOTHETICAL 19.8 KD P	5.24e+01
33	48	54.5	180	4	O99942 HYPOTHETICAL 19.9 KD P	5.24e+01
34	48	54.5	337	10	O92753 COLLAGEN-LIKE (FRAGMEN	5.24e+01
35	48	54.5	396	5	O22651 SIMILAR TO PROCOLLAGEN	5.24e+01
36	48	54.5	432	2	O9XA76 PUTATIVE AMINOPEPTIDAS	5.24e+01
37	48	54.5	493	6	Q28617 ESP-2 (FRAGMENT).	5.24e+01
38	48	54.5	620	14	Q98329 MC163R.	5.24e+01
39	48	54.5	648	5	O09513 HYPOTHETICAL 72.6 KD P	5.24e+01
40	48	54.5	867	14	O69529 HERPESVIRUS 6 IMMEDIAT	5.24e+01
41	48	54.5	997	4	O9Y4G1 KIAA0360 PROTEIN (FRAG	5.24e+01
42	48	54.5	1019	4	Q9Y467 ZINC FINGER PROTEIN HS	5.24e+01
43	48	54.5	1041	4	O9Y441 FILAMIN (FRAGMENT).	5.24e+01
44	48	54.5	1221	11	Q9WU12 FIBULIN-2.	5.24e+01
45	48	54.5	3014	14	O39928 POLYPROTEIN.	5.24e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	53 AA.
ID	Q92728			
AC	Q92728			
DT	01-FEB-1997 (TREMBLrel. 02, Created)			
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)			
DT	01-NOV-1998 (TREMBLrel. 08, Last annotation update)			
DE	RETINOBLASTOMA SUSCEPTIBILITY PROTEIN (RB).			
GN	Rbl.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Homnidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 88320373.			
RA	LEE E.Y., BOOKSTEIN R., YOUNG L.J., LIN C.J., ROSENFELD M.G.,			
RA	LEE W.H.;			
RT	"Molecular mechanism of retinoblastoma gene inactivation in			
RT	retinoblastoma cell line Y79."			
RL	Proc. Natl. Acad. Sci. U.S.A. 85:6017-6021(1988).			
DR	EMBL; M19701; AAA60259.1; -.			
SQ	SEQUENCE 53 AA; 5392 MW; 2D5C3E39 CRC32;			

Query Match 69.3%; Score 61; DB 4; Length 53;
Best Local Similarity 69.2%; Pred. No. 1.38e+01;
Matches 9; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Db	33	PEODSGPEPLV 45
QY	302	AEQESGPPEE-PLV 313

RESULT	2	PRELIMINARY;	PRT;	649 AA.
ID	O94538			
AC	O94538			
DT	01-MAY-1999 (TREMBLrel. 10, Created)			
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)			
DT	01-MAY-1999 (TREMBLrel. 10, Last annotation update)			
DE	HYPOTHETICAL 74.7 KD PROTEIN.			
GN	SPAC167.03C.			
OS	Schizosaccharomyces pombe (Fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Archiascomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomycetes.			
RN	[1]			


```
ID O18563 PRELIMINARY; PRT; 128 AA.
AC O18563;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)
DE IGG AND IGE IMMUNOREACTIVE ANTIGEN RECOGNIZED BY SERA FROM PATIENTS
DE WITH STRONGYLOIDIASIS (FRAGMENT).
OS Strongyloides stercoralis.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Strongyloidea; Strongyloides.
RN [1]
RP SEQUENCE FROM N.A.
RA RAMACHANDRAN S., THOMPSON R.W., GAM A.A., NEVA F.A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U90353; AAB65141.1; -
DR HSSP; P80220; 1DIP. 1
FT NON_TER 1
SQ SEQUENCE 128 AA; 13497 MW; FFBEID10 CRC32;

Query Match 60.2%; Score 53; DB 5; Length 128;
Best Local Similarity 66.7%; Pred. No. 5.92e+00;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 64 AGPEEPGPEEP 75
QY 300 AGAEQESGPEEP 311
||:|: |||||

RESULT 8 PRELIMINARY; PRT; 144 AA.
ID O18564;
AC O18564;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMELrel. 07, Last annotation update)
DE IGG AND IGE IMMUNOREACTIVE ANTIGEN RECOGNIZED BY SERA FROM PATIENTS
DE WITH STRONGYLOIDIASIS (FRAGMENT).
OS Strongyloides stercoralis.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Strongyloidea; Strongyloides.
RN [1]
RP SEQUENCE FROM N.A.
RA RAMACHANDRAN S., THOMPSON R.W., GAM A.A., NEVA F.A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U90354; AAB65142.1; -
DR NON_TER 1
SQ SEQUENCE 144 AA; 15331 MW; D94B8867 CRC32;

Query Match 60.2%; Score 53; DB 5; Length 144;
Best Local Similarity 66.7%; Pred. No. 5.92e+00;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 80 AGPEEPGPEEP 91
QY 300 AGAEQESGPEEP 311
||:|: |||||

RESULT 9 PRELIMINARY; PRT; 152 AA.
ID O18561;
AC O18561;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMELrel. 07, Last annotation update)
DE IGG AND IGE IMMUNOREACTIVE ANTIGEN RECOGNIZED BY SERA FROM PATIENTS
DE WITH STRONGYLOIDIASIS (FRAGMENT).
OS Strongyloides stercoralis.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Strongyloidea; Strongyloides.
RN [1]
RP SEQUENCE FROM N.A.
RA RAMACHANDRAN S., THOMPSON R.W., GAM A.A., NEVA F.A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U90351; AAB65139.1; -
DR NON_TER 1
```

```
SQ SEQUENCE 152 AA; 16546 MW; 3416BBD0 CRC32;

Query Match 60.2%; Score 53; DB 5; Length 152;
Best Local Similarity 66.7%; Pred. No. 5.92e+00;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 118 AGPEEPGPEEP 129
QY 300 AGAEQESGPEEP 311
||:|: |||||

RESULT 10 PRELIMINARY; PRT; 156 AA.
ID O44394;
AC O44394;
DT 01-JUN-1998 (TREMELrel. 06, Created)
DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)
DT 01-AUG-1998 (TREMELrel. 07, Last annotation update)
DE IGG IMMUNOREACTIVE ANTIGEN (FRAGMENT).
OS Strongyloides stercoralis.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Strongyloidea; Strongyloides.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SOUTH-EAST ASIAN;
RA RAMACHANDRAN S., THOMPSON W., NEVA F.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035657; AAB97359.1; -
FT NON_TER 1
SQ SEQUENCE 156 AA; 17339 MW; C408F4F6 CRC32;

Query Match 60.2%; Score 53; DB 5; Length 156;
Best Local Similarity 66.7%; Pred. No. 5.92e+00;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 122 AGPEEPGPEEP 133
QY 300 AGAEQESGPEEP 311
||:|: |||||

RESULT 11 PRELIMINARY; PRT; 317 AA.
ID Q52327;
AC Q52327;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-MAY-1999 (TREMELrel. 10, Last annotation update)
DE TELLURITE RESISTANCE.
GN TELB.
OS Klebsiella aerogenes.
OG Plasmid RK27e-r.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91123183.
RA WALTER E.G., THOMAS C.M., IBBOTSON J.P., TAYLOR D.E.;
RT "Transcriptional analysis, translational analysis, and sequence of the
RT kila-tellurite resistance region of plasmid RK27e-r.";
RL J. Bacteriol. 173:1111-1119(1991).
KW EMBL; M62846; AAA98155.1; -
RW Plasmid.
SQ SEQUENCE 317 AA; 32396 MW; 9A178E47 CRC32;

Query Match 60.2%; Score 53; DB 2; Length 317;
Best Local Similarity 45.3%; Pred. No. 5.92e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 153 GGDHDAGPAEP 163
QY 301 AGAEQESGPEEP 311
||:|: |||||

RESULT 12 PRELIMINARY; PRT; 329 AA.
ID Q9Y5U2
```

```
AC O9Y5U2;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE TUMOR SUPPRESSING STF CDNA 4.
GN TSSC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99172079.
RA LEE M.P., BRANDENBURG S., LANDES G.M., ADAMS M., MILLER G.,
RA FEINBERG A.P.;
RT "Two novel genes in the center of the 11p15 imprinted domain escape
RT genomic imprinting.";
RL Hum. Mol. Genet. 8:683-690(1999).
DR EMBL; AF125568; AAC23579.1; -.
SQ SEQUENCE 329 AA; 34358 MW; 45AF5D55 CRC32;

Query Match 60.2%; Score 53; DB 4; Length 329;
Best Local Similarity 60.0%; Pred. No. 5.92e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 56 GEEDSGPDEP 65
QY 302 AEQESGPEEP 311

RESULT 13
ID O60509 PRELIMINARY; PRT; 365 AA.
AC O60509;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE NSP-LIKE 1 (RTN2-B) (FRAGMENT).
GN NSPL1 OR RTN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98191746.
RA GEISLER J.G., STUBBS L.J., WASSERMAN W.W., MUCENSKI M.L.;
RT "Molecular cloning of a novel mouse gene with predominant muscle and
RT neural expression.";
RL Mamm. Genome 9:274-282(1998).
RN [2]
RP SEQUENCE OF 62-365 FROM N.A.
RX MEDLINE; 98360096.
RA ROEBROEK A.J.M., CONTRERAS B., PAULI I.G.L., VAN DE VEN W.J.M.;
RT "CDNA cloning, genomic organization, and expression of the human RTN2
RT gene, a member of a gene family encoding reticulons.";
RL Genomics 51:98-106(1998).
DR EMBL; AF038540; AAC14910.1; -.
DR EMBL; AF004223; AAC32543.1; -.
FT NON_TER
SQ SEQUENCE 365 AA; 39707 MW; 523E9B57 CRC32;

Query Match 59.1%; Score 52; DB 4; Length 365;
Best Local Similarity 63.6%; Pred. No. 9.26e+00;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 119 DSNSGPPEPL 129
QY 303 EQESGPEEP 313

RESULT 14
ID O75298 PRELIMINARY; PRT; 545 AA.
AC O75298;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
```

```
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE RTN2-A (RTN2-C).
GN RTN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98360096.
RA ROEBROEK A.J.M., CONTRERAS B., PAULI I.G.L., VAN DE VEN W.J.M.;
RT "CDNA cloning, genomic organization, and expression of the human RTN2
RT gene, a member of a gene family encoding reticulons.";
RL Genomics 51:98-106(1998).
DR EMBL; AF004222; AAC32542.1; -.
DR EMBL; AF004224; AAC32544.1; -.
SQ SEQUENCE 545 AA; 59263 MW; 3D7AAB26 CRC32;

Query Match 59.1%; Score 52; DB 4; Length 545;
Best Local Similarity 63.6%; Pred. No. 9.26e+00;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 226 DSNSGPPEPL 236
QY 303 EQESGPEEP 313

RESULT 15
ID O9XIX5 PRELIMINARY; PRT; 650 AA.
AC O9XIX5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE SIMILAR TO EPSTEIN-BARR VIRUS.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Oryza.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. NIPPONBARE;
RA SASAKI T., NAGAMURA Y., YAMAMOTO K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
RT clone:P0681f10.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026295; BAA81856.1; -.
SQ SEQUENCE 650 AA; 67739 MW; 72ACF8C9 CRC32;

Query Match 59.1%; Score 52; DB 10; Length 650;
Best Local Similarity 58.3%; Pred. No. 9.26e+00;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 415 AGAEDEGGDDDP 426
QY 300 AGAEQESGPEEP 311

Search completed: Wed May 10 14:04:54 2000
Job time : 242 secs.
```

W P S R E H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 14:13:54 2000; Maspar time 3.38 seconds
Tabular output not generated. 182.158 Million cell updates/sec

Title: >US-09-376-430-2
Description: (317-342) from US09376430A.pep (23 of 25)
Perfect Score: 164
Sequence: 1 AKTEASPRMLDPQTEKEASGGSLQ 26

Scoring table: PAM 150
Gap 11

Searched: 189863 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
l:geneseqp

Statistics: Mean 20.254; Variance 71.387; scale 0.284

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	62	37.8	464	CD2 associated intrace	6.02e+01
2	62	37.8	464	CD2 associated intrace	6.02e+01
3	62	37.8	464	CD2 associated intrace	6.02e+01
4	62	37.8	553	CD2 associated intrace	6.02e+01
5	62	37.8	553	CD2 associated intrace	6.02e+01
6	62	37.8	553	CD2 associated intrace	6.02e+01
7	62	37.8	553	CD2 associated intrace	6.02e+01
8	62	37.8	553	CD2 associated intrace	6.02e+01
9	62	37.8	553	CD2 associated intrace	6.02e+01
10	61	37.2	208	Yeast MSH1 protein. wit	6.02e+01
11	61	37.2	208	Human TNRL3 protein.	7.44e+01
12	61	37.2	249	TNF related endotheliu	7.44e+01
13	61	37.2	284	Homo sapiens tumour ne	7.44e+01
14	59	36.0	470	Human gravin polypepti	9.18e+01
15	59	36.0	470	Human TIE ligand NL8 p	1.13e+02
16	58	35.4	1584	GAP protein Ital.	1.13e+02
17	58	34.1	160	Human BAI1 protein.	1.39e+02
18	56	34.1	318	Human microtubule-asso	2.11e+02
19	56	34.1	397	NF-YA.	2.11e+02
20	56	34.1	430	Homo sapiens BP202_3 c	2.11e+02
21	56	34.1	430	C.glutamincum isocitrat	2.11e+02
22	56	34.1	763	Streptococcus pneumoni	2.11e+02
23	56	34.1	1026	Human protocadherin-42	2.11e+02
24	56	34.1	1026	Protocadherin clone 42	2.11e+02

24	56	34.1	1132	1	R77417	Human cell cycle prote	2.11e+02
25	56	34.1	1203	1	R58911	Product of alternative	2.11e+02
26	56	34.1	1203	1	R87152	Alternatively spliced	2.11e+02
27	55	33.5	49	1	Y03706	M. tuberculosis LHP po	2.58e+02
28	55	33.5	80	1	W32454	M. tuberculosis tubercul	2.58e+02
29	55	33.5	80	1	W32386	Mycobacterium tubercul	2.58e+02
30	55	33.5	80	1	W64340	Mycobacterium tubercul	2.58e+02
31	55	33.5	95	1	W32444	Mycobacterium tubercul	2.58e+02
32	55	33.5	95	1	W32376	Mycobacterium tubercul	2.58e+02
33	55	33.5	95	1	W81747	M. tuberculosis immuno	2.58e+02
34	55	33.5	95	1	W64321	Mycobacterium tubercul	2.58e+02
35	55	33.5	100	1	Y03705	M. tuberculosis LHP po	2.58e+02
36	55	33.5	100	1	W81706	M. tuberculosis immuno	2.58e+02
37	55	33.5	100	1	W64339	Mycobacterium tubercul	2.58e+02
38	55	33.5	423	1	R89408	Stress activated prote	2.58e+02
39	55	33.5	423	1	R89409	Stress activated prote	2.58e+02
40	55	33.5	424	1	R70768	JNK2 Oncoprotein polyp	2.58e+02
41	55	33.5	515	1	R27704	Opessum kidney pTH/PTH	2.58e+02
42	55	33.5	614	1	Y07038	Breast cancer associat	2.58e+02
43	55	33.5	801	1	W26733	Staphylococcus carnosu	2.58e+02
44	55	33.5	802	1	W64379	Mycobacterium antigen	2.58e+02
45	55	33.5	802	1	W81746	M. tuberculosis fusion	2.58e+02

ALIGNMENTS

RESULT 1
ID W80420 standard; Protein: 464 AA.
AC W80420;
DE 25-JAN-1999 (first entry)
DT CD2 associated intracellular protein encoded by clone LS02-21.
KW CD2-associated intracellular protein; CAIP; human;
KW Lymphocyte proliferation; interleukin-2; immunological disease;
KW psoriasis; transplant rejection; T cell proliferation;
KW immunodeficiency; cancer; chemotherapy.
OS Homo sapiens.
PN US5837844-A.
PD 17-NOV-1998.
PF 07-JUN-1995; 484709.
PR 07-JUN-1995; US-484709. \uparrow
PA (BIOU) BIOGEN INC.
PI Hsu Y;
DR WPI; 99-023540/02.
DR N-PSDB; V63343.
PT Nucleic acid encoding CD2-associated intracellular protein - for
PT recombinant production of the protein, used to regulate T cell
PT proliferation, e.g. for treating psoriasis, transplant rejection and
PT immunodeficiency
PS Claim 1; Fig 4; 33pp; English.
CC The present sequence is encoded by the LS02-21 clone, and represents a
CC CD2-associated intracellular protein (CAIP). The DNA sequence is
CC isolated from human B cell cDNA, and the encoded protein binds to
CC the intracellular domain of CD2. The CAIP proteins are used as
CC agonists or antagonists of natural CAIP. Antagonists inhibit
CC lymphocyte proliferation (they reduce production of interleukin-2)
CC and are used to treat immunological diseases such as psoriasis and
CC transplant rejection associated with T cell proliferation. Agonists
CC are used to treat immunodeficiency, particularly when caused by
CC immunosuppressant treatments or cancer chemotherapy.
SQ Sequence 464 AA;

Query Match 37.8%; Score 62; DB 1; Length 464;
Best Local Similarity 55.6%; Pred. No. 6.02e+01;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 346 KIPPERPEMLPNTKEE 363
QY 318 KTEASPRMLDPQTEKE 335

RESULT 2
ID W26496 standard; Protein: 464 AA.
AC W26496;

US-09-376-430-2-23.rag

Thu May 11 06:50:16 2000

```

05-JAN-1998 (first entry)
CD2 associated intracellular protein CAIP LS02-21.
CAIP; CD2 associated intracellular protein; lymphocyte;
signal transduction; therapy; diagnosis; antibody; agonist.
Homo sapiens. Location/Qualifiers
Key 1..5 /note= "primer-encoded peptide"
Peptide 6..37
Domain /label= N-terminal
Domain 38..464 /label= Central
Binding_site 88..147 /label= SH3-2
Binding_site 258..317 /label= SH3-3
Binding_site 318..329 /label= HP-1
Binding_site 348..355 /label= HP-2
Binding_site 382..389 /label= HP-3
Binding_site 401..411 /label= HP-4
US5641748-A.
12-AUG-1997. 484710.
07-JUN-1995; US-484710.
07-JUN-1995; US-484710.
(BIOJ ) BIOGEN INC.
Hsu Y.
WPI: 97-414582/38.
N-PSDB; T87046.
Antibodies to CD2-associated intracellular protein - useful for
research, diagnosis, therapy monitoring, etc.
Claim 1; Fig 4; 33pp; English.
This polypeptide comprises a novel intracellular lymphocyte
protein, termed CD2 associated intracellular domain of CD2. Its sequence
can interact with the intracellular domain of CD2. Its sequence
was deduced from a cDNA clone (877046) isolated from a B lymphocyte
library. CAIP polypeptides (see also W26495 and W26497) possess
common central domains and unique N-terminal regions. They are
useful for: identifying cells which preferentially express CAIP
e.g. lymphocytes or testis cells; the production of peptides that
modulate lymphocyte activation in vivo or in vitro; for analysis of
lymphocyte, e.g. CD2 mediated, activation; to generate anti-CAIP
antibodies; and for producing CAIP binding fragments of CD2, which
can be used in vitro or in vivo to modulate lymphocyte activation.
Antibodies which bind to CAIP are claimed and can be used to detect
CAIP polypeptides in a tissue or bodily fluid sample. CAIP agonists
can be used to promote restoration of a compromised immune system.
Antagonists can be used to inhibit T-cell proliferation e.g. for
treatment of psoriasis or transplant tissue rejection.
Sequence 464 AA;

Query Match 37.8%; Score 62; DB 1; Length 464;
Best Local Similarity 55.6%; Pred. No. 6.02e+01;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 346 KIPPEREMLPNRTEEKE 363
QY 318 KTEASPRMLDPQTEEKE 335
| : | | | : | | | |
RESULT 3
ID W25116 standard; Protein; 464 AA.
AC W25116;
DE 17-NOV-1997 (first entry)
DE CD2-associated intracellular protein encoded by cDNA clone LS02-21.
DE CD2-associated intracellular protein; CAIP; SH3 domain; binding;
KW lymphocyte proliferation inhibitor; autoimmune disease; psoriasis;
KW organ transplant rejection; T cell.
OS Homo sapiens. Location/Qualifiers
FH Key Location/Qualifiers
FT Misc_difference 26 /note= "not specified"
FT Misc_difference 40 /note= "not specified"
FT Misc_difference 64 /note= "not specified"
FT Misc_difference 65 /note= "not specified"
FT Misc_difference 107 /note= "not specified"
FT US5837844-A.
PN 17-NOV-1998.
PF 07-JUN-1995; 484709.

05-JAN-1998 (first entry)
CD2 associated intracellular protein CAIP LS02-21.
CAIP; CD2 associated intracellular protein; lymphocyte;
signal transduction; therapy; diagnosis; antibody; agonist.
Homo sapiens. Location/Qualifiers
Key 1..5 /note= "primer-encoded peptide"
Peptide 6..37
Domain /label= N-terminal
Domain 38..464 /label= Central
Binding_site 88..147 /label= SH3-2
Binding_site 258..317 /label= SH3-3
Binding_site 318..329 /label= HP-1
Binding_site 348..355 /label= HP-2
Binding_site 382..389 /label= HP-3
Binding_site 401..411 /label= HP-4
US5641748-A.
12-AUG-1997. 484710.
07-JUN-1995; US-484710.
07-JUN-1995; US-475894.
(BIOJ ) BIOGEN INC.
Hsu Y.
WPI: 97-340971/31.
N-PSDB; T79834.
New CD2-associated intracellular proteins (CAIPs) - are modulators
of CD2 mediated signalling, useful for modulating T cell
proliferation, e.g. for treating immunodeficiency, psoriasis and
transplant rejection
Claim 5; Column 41-44; 33pp; English.
W25116 gives the sequence of a human CD2-associated intracellular
protein (CAIP). The CAIP polypeptide may bind to naturally occurring
CAIP inhibiting binding to CD2 SH3 domains; inhibit CD2 directly or
inhibit CAIP mediated intracellular signalling, ultimately inhibiting
lymphocyte proliferation. The polypeptide can be used to treat T cell
mediated immune disorders such as psoriasis and organ transplant
rejection. Immunogens containing the polypeptide are used to raise
antibodies (Ab) reactive with CAIP epitopes, for detection of
CAIP-expressing cells.
Sequence 464 AA;

Query Match 37.8%; Score 62; DB 1; Length 464;
Best Local Similarity 55.6%; Pred. No. 6.02e+01;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 346 KIPPEREMLPNRTEEKE 363
QY 318 KTEASPRMLDPQTEEKE 335
| : | | | : | | | |
RESULT 4
ID W80419 standard; Protein; 553 AA.
AC W80419;
DE 25-JAN-1999 (first entry)
DE CD2 associated intracellular protein encoded by clone YM06.
DE CD2-associated intracellular protein; CAIP; human;
KW lymphocyte proliferation; interleukin-2; immunological disease;
KW psoriasis; transplant rejection; T cell proliferation;
KW immunodeficiency; cancer chemotherapy.
OS Homo sapiens. Location/Qualifiers
FH Key Location/Qualifiers
FT Misc_difference 26 /note= "not specified"
FT Misc_difference 40 /note= "not specified"
FT Misc_difference 64 /note= "not specified"
FT Misc_difference 65 /note= "not specified"
FT Misc_difference 107 /note= "not specified"
FT US5837844-A.
PN 17-NOV-1998.
PF 07-JUN-1995; 484709.

```

PR 07-JUN-1995; US-484709.
PA (BIOJ) BIOGEN INC.
PI Hsu Y;
DR WPI: 99-023540/02.
DR N-PSDB; V63342.
PT Nucleic acid encoding CD2-associated intracellular protein - for
PT recombinant production of the protein, used to regulate T cell
PT proliferation, e.g. for treating psoriasis, transplant rejection and
PT immunodeficiency
PS Claim 1; Fig 2; 33pp; English.
CC The present sequence is encoded by the YM06 clone, and represents a
CC CD2-associated intracellular protein (CAIP). The DNA sequence is
CC isolated from human B cell cDNA, and the encoded protein binds to
CC the intracellular domain of CD2. The CAIP proteins are used as
CC agonists or antagonists of natural CAIP. Antagonists inhibit
CC lymphocyte proliferation (they reduce production of interleukin-2)
CC and are used to treat immunological diseases such as psoriasis and
CC transplant rejection associated with T cell proliferation. Agonists
CC are used to treat immunodeficiency, particularly when caused by
CC immunosuppressant treatments or cancer chemotherapy.
SQ Sequence 553 AA;

Query Match 37.8%; Score 62; DB 1; Length 553;

Best Local Similarity 55.6%; Pred. No. 6.02e+01;

Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 435 KIPPERPEMLPNRTEEKE 452
QY 318 KTEAESPRMLDPQTEEKE 335
| : | | | : | | | |

RESULT 5
ID W26495 standard; Protein; 553 AA.
AC W26495;
DT 05-JAN-1998 (first entry)
DE CD2 associated intracellular protein CAIP (YM06).
KW CAIP; CD2 associated intracellular protein; lymphocyte;
KW signal transduction; therapy; diagnosis; antibody; agonist.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 26
FT Misc_difference 40 /note= "encoded by ANT"
FT Misc_difference 64 /note= "encoded by TAA"
FT Misc_difference 64 /note= "encoded by TAA"
FT Misc_difference 65 /note= "encoded by TAG"
FT Misc_difference 90 /note= "encoded by CTN"
FT Misc_difference 107 /note= "encoded by NGC"
FT Domain 1..85
FT Domain 86..553
FT Binding_site 66..126
FT Binding_site 177..236
FT Binding_site 347..406
FT Binding_site 408..415
FT Binding_site 437..445
FT Binding_site 471..478
FT Binding_site 490..497
FT Binding_site /label= HP-4
PN US5656438-A.
PD 12-AUG-1997.
PF 07-JUN-1995; 484710.

PR 07-JUN-1995; US-484710.
PA (BIOJ) BIOGEN INC.
PI Hsu Y;
DR WPI: 97-414582/38.
DR N-PSDB; T87045.
PT Antibodies to CD2-associated intracellular protein - useful for
PT research, diagnosis, therapy monitoring, etc.
PS Claim 1; Fig 2; 33pp; English.
CC This polypeptide comprises a novel intracellular lymphocyte
CC protein, termed CD2 associated intracellular protein (CAIP), that
CC can interact with the intracellular domain of CD2. Its sequence
CC was deduced from a cDNA clone (T87045) isolated from a B lymphocyte
CC library. CAIP polypeptides (see also W26496 and W26497) possess
CC common central domains and unique N-terminal regions. They are
CC useful for: identifying cells which preferentially express CAIP
CC e.g. lymphocytes or testis cells; the production of peptides that
CC modulate lymphocyte activation in vivo or in vitro; for analysis of
CC lymphocytes, e.g. CD2 mediated, activation; to generate anti-CAIP
CC antibodies; and for producing CAIP binding fragments of CD2, which
CC can be used in vitro or in vivo to modulate lymphocyte activation.
CC Antibodies which bind to CAIP are claimed and can be used to detect
CC CAIP polypeptides in a tissue or bodily fluid sample. CAIP agonists
CC can be used to promote restoration of a compromised immune system.
CC Antagonists can be used to inhibit T-cell proliferation e.g. for
CC treatment of psoriasis or transplant tissue rejection.
SQ Sequence 553 AA;

Query Match 37.8%; Score 62; DB 1; Length 553;

Best Local Similarity 55.6%; Pred. No. 6.02e+01;

Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 435 KIPPERPEMLPNRTEEKE 452
QY 318 KTEAESPRMLDPQTEEKE 335
| : | | | : | | | |

RESULT 6
ID W25115 standard; Protein; 553 AA.
AC W25115;
DT 17-NOV-1997 (first entry)
DE CD2-associated intracellular protein encoded by cDNA clone YM06.
KW CD2-associated intracellular protein; CAIP; SH3 domain; binding;
KW lymphocyte proliferation inhibitor; autoimmune disease; psoriasis;
KW organ transplant rejection; T cell.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 26 /note= "X= undefined in the specification"
FT Misc_difference 40 /note= "X= undefined in the specification"
FT Misc_difference 64..65 /note= "X= undefined in the specification"
FT domain 66..126
FT Misc_difference 107 /label= SH3-1_domain
FT domain 176..236 /note= "X= undefined in the specification"
FT domain 346..406 /label= SH3-2_domain
FT domain 407..415 /label= SH3-3_domain
FT domain 437..445 /note= "SH3 binding domain"
FT domain 471..478 /label= HP-2
FT domain 490..497 /note= "SH3 binding domain"
FT domain 490..497 /label= HP-4
FT domain /note= "SH3 binding domain"
PN US5641748-A.

PD 24-JUN-1997.
 PF 07-JUN-1995; 475894.
 PR 07-JUN-1995; US-475894.
 PA (BIOJ) BIOGEN INC.
 PI Hsu Y.
 DR WPI: 97-340971/31.
 DR N-PSDB; T79833.
 PT New CD2-associated intracellular proteins (CAIPs) - are modulators
 of CD2 mediated signalling, useful for modulating T cell
 proliferation, e.g. for treating immunodeficiency, psoriasis and
 transplant rejection
 PS Claim 1; Column 37-40; 33pp; English.
 CC W3115 gives the sequence of a human CD2-associated intracellular
 protein (CAIP). The CAIP polypeptide may bind to naturally occurring
 CAIP inhibiting binding to CD2 SH3 domains; inhibit CD2 directly or
 inhibit CAIP mediated intracellular signalling, ultimately inhibiting
 lymphocyte proliferation. The polypeptide can be used to treat T cell
 mediated immune disorders such as psoriasis and organ transplant
 rejection. Immunogens containing the polypeptide are used to raise
 antibodies (Ab) reactive with CAIP epitopes, for detection of
 CAIP-expressing cells.
 CC Sequence 553 AA;
 SQ

Query Match 37.8%; Score 62; DB 1; Length 553;
 Best Local Similarity 55.6%; Pred. No. 6.02e+01;
 Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 435 KIPPERPEMLPNRTTEEKE 452
 | : | | | : | | | |
 QY 318 KTEASPRMLDPQTEEKE 335

RESULT 7
 ID W37724 standard; Protein; 659 AA.
 AC W37724;
 DT 09-JUN-1998 (first entry)
 DE CD2 associated intracellular protein (CAIP).
 KW Lymphocyte protein; CD2 associated intracellular protein; CAIP;
 KW antagonistic; inhibition; CAIP mediated intracellular signalling;
 KW lymphocyte proliferation; psoriasis; agonist; chemotherapy; cancer.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Domain 1..47 "SH3-1"
 /note= "SH3-1"
 FT Domain 95..0
 /note= "SH3-2"
 FT Domain 264..317
 /note= "SH3-3"
 FT Binding_site 329..337
 /note= "SH3 binding domain 1"
 FT Binding_site 358..367
 /note= "SH3 binding domain 2"
 FT Binding_site 392
 /note= "SH3 binding domain 3"
 FT Binding_site 411..419
 /note= "SH3 binding domain 4"
 FT misc_difference 12
 /note= "encoded by CTN"
 FT misc_difference 29
 /note= "encoded by NGC"
 PN W09749808-A1.
 PD 31-DEC-1997.
 PF 27-JUN-1996; U11042.
 PR 27-JUN-1996; WO-U11042.
 PA (BIOJ) BIOGEN INC.
 PI Hsu Y.
 DR WPI: 98-077172/07.
 DR N-PSDB; V18892.
 PT New isolated CD2 associated intracellular protein - used to develop
 products for use as antagonists, e.g. for inhibiting lymphocyte
 proliferation, or agonists, e.g. for treating immunodeficiency
 disorders
 PS Claim 2; Pages 43-45; 55pp; English.

CC This is the novel lymphocyte protein, CD2 associated intracellular
 protein (CAIP). The CAIP polypeptides can have antagonistic activity,
 and are capable of inhibiting CD2/CAIP binding, inhibiting CD2 or CAIP
 mediated intracellular signalling, or inhibiting lymphocyte
 proliferation. It can be used to inhibit an immune disorder
 characterised by unwanted T cell proliferation, e.g. as in psoriasis or
 unwanted rejection of transplant tissue. They can also have agonist
 activity and can be used to treat disorders such as immunodeficiency
 characterised by an insufficient lymphocyte of T cell activity, such as
 after chemotherapy for treating cancer.
 CC Sequence 659 AA;
 SQ

Query Match 37.8%; Score 62; DB 1; Length 659;
 Best Local Similarity 55.6%; Pred. No. 6.02e+01;
 Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 357 KIPPERPEMLPNRTTEEKE 374
 | : | | | : | | | |
 QY 318 KTEASPRMLDPQTEEKE 335

RESULT 8
 ID R76066 standard; Protein; 959 AA.
 AC R76066;
 DT 16-JAN-1996 (first entry)
 DE Yeast MSH1 protein.
 KW Mismatch repair; MSH2; primer; identification; defect; alteration;
 KW cancer; tumour; vaccine.
 OS Saccharomyces cerevisiae.
 PN W09514085-A2.
 PD 26-MAY-1995.
 PF 17-NOV-1994; U13385.
 PR 17-NOV-1993; US-154792.
 PR 07-DEC-1993; US-163449.
 PR 13-JUN-1994; US-259310.
 PA (UIVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
 PA (DAND) DANA FARBER CANCER INST.
 PI Fishel R, Kolodner RD, Reenan RAG;
 DR WPI: 95-200377/26.
 DR N-PSDB; Q93913.
 PT Determining alteration in human mismatch repair pathways - used in
 the diagnosis, prognosis and therapy of cancers and in screening
 PT assays
 PS Example 15; Page 149-153; 256pp; English.
 CC R76066 is the S.cerevisiae mismatch repair pathway protein MSH1. It
 is used in the analysis of possible mutations in the human mismatch
 repair genes. Defects or alterations in such a gene result in the
 accumulation of unstable repeated DNA sequences, a feature of a number
 of different cancers. The identification of a defect in the mismatch
 repair pathway can be diagnostic of a predisposition to cancer and
 prognostic for a particular mammalian cancer e.g colorectal, ovarian,
 endometrial (uterine), renal, bladder, skin, rectal and bowel. The
 nucleotide sequences and polypeptides of the hMSH2 gene may also be
 used for therapy and in vaccines.
 CC Sequence 959 AA;
 SQ

Query Match 37.8%; Score 62; DB 1; Length 959;
 Best Local Similarity 42.9%; Pred. No. 6.02e+01;
 Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Db 529 AKQNAADVTRMLDIDVKKES 549
 | : | | | : | | | |
 QY 317 AKTEASPRMLDPQTEEKEAS 337

RESULT 9
 ID R76070 standard; Protein; 971 AA.
 AC R76070;
 DT 16-JAN-1996 (first entry)
 DE Yeast MSH1 protein with I2CA5 epitope tag.
 KW Mismatch repair; MSH2; primer; identification; defect; alteration;
 KW cancer; tumour; vaccine.
 OS Saccharomyces cerevisiae.

PN W09514085-A2.
 PD 26-MAY-1995.
 PF 17-NOV-1994; U13385.
 PR 17-NOV-1993; US-154792.
 PR 07-DEC-1993; US-163449.
 PR 13-JUN-1994; US-259310.
 PA (UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
 PA (DAND) DANA FARBER CANCER INST.
 PI Fishel R, Kolodner RD, Reenan RAG;
 PI WPI; 95-200377/26.
 PT Determining alteration in human mismatch repair pathways - used in
 PT the diagnosis, prognosis and therapy of cancers and in screening
 PT assays
 PS Disclosure; Page 167-171; 256pp; English.
 CC R76070 is the S.cerevisiae mismatch repair pathway protein MSH1,
 CC containing the I2CA5 epitope. It is used in the analysis of possible
 CC mutations in the human mismatch repair genes. Defects or alterations
 CC in such a gene result in the accumulation of unstable repeated DNA
 CC sequences, a feature of a number of different cancers. The
 CC identification of a defect in the mismatch repair pathway can be
 CC diagnostic of a predisposition to cancer and prognostic for a
 CC particular mammalian cancer e.g colorectal, ovarian, endometrial
 CC (uterine), renal, bladder, skin, rectal and bowel. The nucleotide
 CC sequences and polypeptides of the hMSH2 gene may also be used for
 CC therapy and in vaccines.
 SQ Sequence 971 AA;

Query Match 37.8%; Score 62; DB 1; Length 971;
 Best Local Similarity 42.9%; Pred. No. 6.02e+01;
 Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Db 541 AKQNAADVTRMLDIDVKDKKES 561
 QY 317 AKTEASPRMLDPQTEKEAS 337
 |||: ||||: |||

RESULT 10
 ID W93590 standard; Protein; 208 AA.
 AC W93590;
 DT 18-JUN-1999 (first entry)
 DE Human TNFR3 protein.
 KW Tumour necrosis factor receptor; signal transducer molecule; TNF; APO4;
 KW developmental abnormality; gestational abnormality; prostate cancer;
 KW APO6; APO8; APO9; TNRL-1; TNRL-3; diagnosis; treatment; therapy; disease;
 KW cytoplasmic domain; immunogen; antibody preparation; breast carcinoma;
 KW apoptosis; human.
 OS Homo sapiens.
 PN W0911791-A2.
 PD 11-MAR-1999.
 PF 04-SEP-1998; U18393.
 PR 05-SEP-1997; US-924634.
 PA (UNIW) UNIV WASHINGTON.
 PI Chaudhary PM;
 DR WPI; 99-205191/17.
 DR N-PSDB; X23424.
 PT New Tumor Necrosis Factor family receptor polypeptides and ligands -
 PT useful for diagnosis and treatment of prostate cancer and
 PT developmental or gestational abnormalities
 PS Claim 40; Fig 13A; 156pp; English.
 CC This invention describes isolated Tumor Necrosis Factor (TNF) family
 CC receptor polypeptides: APO4, APO6, APO8 and APO9 or their active
 CC fragments, and isolated TNF related ligands 1 and 3 (TNRL1 and TNRL3) or
 CC their active fragments. APO4 is useful for diagnosing prostate cancer
 CC by determining levels of APO4 in an individual. Prostate cancer can also
 CC be treated using APO4 selective binding agents linked to a therapeutic
 CC moiety. APO4 polypeptides are also useful for identifying selective
 CC binding agents; useful in diagnosis/treatment of disease by binding of
 CC agents to the polypeptide/active fragment which is extracellular, or
 CC expressed on the cell surface. The binding is preferably performed in
 CC vivo. APO4 polypeptides/ active fragments are also useful for screening
 CC for agonists and antagonists by binding and observing the change in APO4
 CC activity. Effective pharmacological agents useful in diagnosis or
 CC treatment of disease are also identified using APO4 polypeptides/active

CC fragments and APO4 signal transducer molecules that specifically interact
 CC with a cytoplasmic domain of APO4 and detecting a change in level of APO4
 CC activity. The method is performed in vivo or in vitro. APO polypeptides
 CC are all useful as immunogens for preparing antibodies. APO4 is also
 CC useful for diagnosis/treatment of developmental or gestational
 CC abnormalities. APO8 was transfected to human breast carcinoma cell line
 CC MCF-7, and induced apoptosis.
 SQ Sequence 208 AA;

Query Match 37.2%; Score 61; DB 1; Length 208;
 Best Local Similarity 40.9%; Pred. No. 7.44e+01;
 Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Db 23 EDQDPSELNPQTEESQDPAPFL 44
 QY 320 EAESPRMLDPQTEKEASGSSL 341
 | : | : ||||| : : |

RESULT 11
 ID W29745 standard; Protein; 249 AA.
 AC W29745;
 DT 27-OCT-1998 (first entry)
 DE TNF related endothelium proliferative agent protein.
 KW TNF; endothelium proliferative agent; TREPA; wound healing; cancer;
 KW tissue grafting; vascularisation; apoptosis; autoimmune; birth control.
 OS Homo sapiens.
 PN W09835061-A2.
 PD 13-AUG-1998.
 PF 12-FEB-1998; U02859.
 PR 10-FEB-1998; US-021706.
 PR 12-FEB-1997; US-798692.
 PA (ABBO) ABBOTT LAB.
 PI Wiley SR;
 DR WPI; 98-447255/38.
 DR N-PSDB; V47613.
 PT Detecting nucleic acid encoding TREPA - useful for diagnosis and
 PT treatment of autoimmune disease, tumours and inflammation
 PS Claim 16; Page 123-4; 142pp; English.
 CC The TNF-related endothelium proliferative agent (TREPA), or its
 CC activators or agonists, are used to treat a deficit of TREPA, e.g. to
 CC promote wound healing or tissue grafting, by promoting vascularisation,
 CC also to induce apoptosis for treating cancer and eliminating autoreactive
 CC T cells, as an adjunct to cancer chemotherapy or antiviral treatment.
 CC TREPA peptides can also be used to target cytotoxic agents or for
 CC affinity isolation of the corresponding receptor, the nucleic acid for
 CC which can be used to transform tumour cells to render them more
 CC responsive to TREPA and to screen for TREPA mimics.
 CC Ribozymes, antisense RNA, antibodies or peptides, are used to treat
 CC TREPA-associated diseases, e.g. tumours and metastases (by inhibiting
 CC vascularisation), inflammation or a wide range of autoimmune conditions,
 CC conditions involving abnormal stimulation of epithelial cells (e.g.
 CC atherosclerosis), for birth control (inhibiting ovulation and placental
 CC formation) or other angiogenic conditions (e.g. ulcers).
 SQ Sequence 249 AA;

Query Match 37.2%; Score 61; DB 1; Length 249;
 Best Local Similarity 40.9%; Pred. No. 7.44e+01;
 Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Db 64 EDQDPSELNPQTEESQDPAPFL 85
 QY 320 EAESPRMLDPQTEKEASGSSL 341
 | : | : ||||| : : |

RESULT 12
 ID W47525 standard; Protein; 284 AA.
 AC W47525;
 DT 21-JUL-1998 (first entry)
 DE Homo sapiens tumour necrosis factor related ligand (TRELL).
 KW TRELL; tumour necrosis factor related ligand; tnfr; treatment;
 KW cancer; autoimmune disease; immune system; stimulation; suppression;
 KW graft rejection.
 OS Homo sapiens.

```
PN WO9805783-A1.
PD 12-FEB-1998.
PF 07-AUG-1997; U13945
PR 18-MAR-1997; US-040820.
PR 07-AUG-1996; US-023541.
PR 18-OCT-1996; US-028515.
PA (BIOJ ) BIOGEN INC.
PA (UVGE-) UNIV GENEVA FACULTY MEDICINE.
PI Browning JL, Chicheportiche Y;
DR WPI: 98-145619/13.
DR N-PSDB: V18600.
PT Tumour necrosis factor related ligand - useful for, e.g. treating
PT cancer, auto-immune disease and immune responses to tissue grafts
PS Claim 12; Pages 50-51; 69pp; English.
CC The sequence is that of human tumour necrosis factor related
CC ligand (TREL). TREL or active fragments can be included with a
CC carrier in pharmaceutical compositions to treat cancer, autoimmune
CC diseases or immune responses to tissue grafts, or to stimulate or
CC suppress the immune system. It is useful to screen for TREL
CC receptors, by labelling with a detectable label and screening
CC compositions for binding. Agents interfering with TREL-receptor
CC binding can also be screened for, can then be administered,
CC optionally with interferon- gamma, to induce cell death or
CC treat, suppress or alter immune responses (especially involving human
CC adenocarcinoma cells) involving a signal pathway between TREL and its
CC receptor. It's coding sequence can be used in gene therapy for
CC TREL-related disorders in mammals (especially humans), e.g. tumours,
CC autoimmune and inflammatory diseases or inherited genetic disorders,
CC by introducing into cells, and expressing, therapeutically effective
CC amounts of a vector, e.g. a virus comprising a gene encoding TREL.
CC It may also be of use in the preparation of prepare probes for
CC screening natural/synthetic DNAs for TREL-encoding sequences
CC and for antisense therapy.
CC Sequence 284 AA;
SQ

Query Match 37.2%; Score 61; DB 1; Length 284;
Best Local Similarity 40.9%; Pred. No. 7.44e+01;
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Db 99 EQDDPSELNPQTEESQDPAPFL 120
QY 320 EAESPRMLDPQTEKEASGGSL 341

RESULT 13
ID W53863 standard; peptide; 1780 AA.
AC W53863.
DE Human gravin polypeptide.
KW Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;
KW CAMP-dependent protein kinase; protein kinase C; autoimmune disease;
KW Myasthenia gravis; nicotinic acetylcholine receptor.
OS Homo sapiens.
PN US5741890-A.
PD 21-APR-1998.
PF 19-DEC-1996; 769309.
PR 19-DEC-1996; US-769309.
PA (UYOR-) UNIV OREGON HEALTH SCI.
PI Klauack TM, Nauert JB, Scott JD;
DR WPI: 98-260552/23.
DR N-PSDB: V23545.
PT New polypeptide fragments of protein kinase binding protein gravin -
PT are useful for the study of modulation of action between gravin and
PT protein kinase(s)
PS Example 1; Column 19-32; 32pp; English.
CC This sequence corresponds to the human gravin polypeptide, and represents
CC a polypeptide of the invention. The polypeptides are fragments capable of
CC binding to type II regulatory subunit of CAMP-dependent protein kinase
CC (PKA). Gravin is a kinase anchoring protein that binds to type II
CC regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an
CC antigen of the autoimmune disease Myasthenia gravis (MG), where a patient
CC develops antibodies against their own nicotinic acetylcholine receptors.
CC The polypeptides are useful for providing analogues of gravin in the

CC study of the modulation (e.g. blocking, inhibiting and stimulating) of
CC interactions between gravin and kinase. The peptides are involved in the
CC modulation of gravin-kinase interactions.
SQ
Sequence 1780 AA;

Query Match 36.6%; Score 60; DB 1; Length 1780;
Best Local Similarity 34.8%; Pred. No. 9.19e+01;
Matches 8; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Db 942 ABEPTPTVTEPLPENREARGDTV 964
QY 319 TEAESPRLMDPQTEKEASGGSL 341

RESULT 14
ID Y05398 standard; Protein; 470 AA.
AC Y05398;
DE Human TIE ligand N18 protein sequence.
KW Human TIE ligand; N18; N14; N15; N18; Ig homology domain; angiogenesis;
KW EGF homology domain; receptor tyrosine kinase; vascular endothelial cell;
KW early haemopoietic cell; haemopoiesis; neovascularisation; wound healing;
KW endothelial cell; growth inhibitor; apoptosis inducer; tumour cell;
KW vasculogenesis; detection; diagnosis; therapy.
OS Homo sapiens.
PN WO9915653-A2.
PD 01-APR-1999.
PF 14-SEP-1998; U19093.
PR 29-OCT-1997; US-960507.
PR 19-SEP-1997; US-933821.
PA (GETH ) GENENTECH INC.
PI Botstein D, Ferrara N, Goddard A, Godowski PJ, Gurney AL,
PI Hillan K, Roy M, Schwall R, Tumas D;
DR WPI: 99-263480/22.
DR N-PSDB: X36343.
DE New isolated TIE ligand homologs for, e.g. developing products for
DE treatment of tumors
PS Claim 8; Fig 7; 132pp; English.
CC This sequence is the human tyrosine kinase containing Ig and EGF
CC homology domains (TIE) ligand of the invention, designated NL8.
CC The TIE receptors are receptor tyrosine kinases which are expressed in
CC vascular endothelial cells and early haemopoietic cells. The TIE
CC receptors are believed to be actively involved in angiogenesis, and may
CC play a role in haemopoiesis as well. The TIE ligand homologs can promote
CC the survival and/or growth and/or differentiation of TIE receptor
CC expressing cells. They can be used for promoting neovascularisation in
CC wound healing and for promoting angiogenic processes, such as for
CC inducing collateral vascularisation in an ischaemic heart or limb, or for
CC promoting bone development and/or maturation and/or growth in a patient
CC or muscle growth and development. The TIE ligand homologs and antibodies
CC can inhibit the growth of endothelial cells and induce apoptosis of
CC cells, particularly tumour cells. They can inhibit vasculogenesis,
CC particularly the vascularisation of tumour cells. The antibodies can also
CC inhibit vascularisation of a cell in which a gene encoding an NL1, NL5,
CC NL8 or NL4 polypeptide is amplified. The products can also be used for
CC detection, diagnosis, drug screening and production of transgenic
CC animals.
SQ
Sequence 470 AA;

Query Match 36.0%; Score 59; DB 1; Length 470;
Best Local Similarity 33.3%; Pred. No. 1.13e+02;
Matches 6; Conservative 8; Mismatches 3; Indels 1; Gaps 1;

Db 214 SDTSRMLDPAPPEPQDQDT 231
QY 321-AESPRMLDPQTE-EKEAS 337

RESULT 15
ID R59923 standard; protein; 2938 AA.
AC R59923;
DR 22-FEB-1995 (first entry)
DE GAP protein Irai.
```

Thu May 11 06:50:16 2000

Ras; GTPase activating protein; GAP; GAP related domain; GRD;
RAS2; v-Ras; heat shock; neurofibromatosis type 1; NFI.
Saccharomyces cerevisiae.
PN WO9416069-A.
PD 21-JUL-1994.
PF 12-JAN-1994; U00198.
PR 15-JAN-1993; US-004824.
PA (SCHE) SCHERING CORP.
PI Kaziro Y, Nakafuku M;
DR WPI: 94-249216/30.
PT Blocking Ras-induced effects on a cell - by introducing a GTPase
PT activating protein to the cell, used esp. in treatment of cancers
PS Disclosure; Page 53-62; 87pp; English.
CC Human neurofibromatosis type 1 (NFI)-GAP related domain (GRD)
CC mutant clones NF201 (given in R59221) and NF204 (R59922) show
CC strong suppression activity for RAS2val19, and inhibit
CC v-Ras-induced transformation in mammalian cells. The mutation
CC sites of these proteins were located in one of the most conserved
CC regions of GRD. These sites were compared with those of other
CC GRD family proteins, yeast Ira2 (R59926) and Ira1 (R59923)
CC human GAP (R59924) and Schizosaccharomyces pombe Gap1 (R59925).
CC Sequence 2938 AA;
SQ

```
Query Match      36.0%; Score 59; DB 1; Length 2938;
Best Local Similarity 20.0%; Pred. No. 1.13e+02;
Matches      5; Conservative 14; Mismatches 6; Indels 0; Gaps 0;
```

```
Db      310 TKTDADTPSTMNTNNNNNNNSANL 334
      :||:|:|:| :: ::: ::::|
QY      317 AKTEAESPRMLDPQTEEKEASGSL 341
```

Search completed: Wed May 10 14:14:01 2000
Job time : 7 secs.

W O R L D
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 14:14:19 2000; Maspar time 47.20 Seconds

Tabular output not generated. 7.137 Million cell updates/sec

Title: >US-09-376-430-2
Description: (317-342) from US09376430A.pap (23 of 25)
Perfect Score: 164
Sequence: 1 AKTEASPRMLDPQTEKEASGSLQ 26

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 19.114; Variance 67.502; scale 0.283

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	62	37.8	464	1	US-08-475-Sequence 4, Applicatio	2.73e+01
2	62	37.8	464	1	US-08-484-Sequence 4, Applicatio	2.73e+01
3	62	37.8	464	2	US-08-484-Sequence 4, Applicatio	2.73e+01
4	62	37.8	553	1	US-08-475-Sequence 2, Applicatio	2.73e+01
5	62	37.8	553	1	US-08-484-Sequence 2, Applicatio	2.73e+01
6	62	37.8	553	2	US-08-484-Sequence 2, Applicatio	2.73e+01
7	60	36.6	1780	1	US-08-769-Sequence 5, Applicatio	4.20e+01
8	59	35.0	189	1	US-08-446-Sequence 1, Applicatio	5.20e+01
9	59	35.0	470	2	US-08-933-Sequence 6, Applicatio	5.20e+01
10	59	36.0	2938	3	PCT-US94-0Sequence 3, Applicatio	5.20e+01
11	56	34.1	368	2	US-08-805-Sequence 3, Applicatio	9.82e+01
12	56	34.1	335	2	US-09-057-Sequence 3, Applicatio	9.82e+01
13	56	34.1	1026	1	US-08-268-Sequence 95, Applicatio	9.82e+01
14	56	34.1	1026	1	US-08-453-Sequence 95, Applicatio	9.82e+01
15	56	34.1	1026	1	US-07-998-Sequence 95, Applicatio	9.82e+01
16	56	34.1	1026	3	PCT-US95-0Sequence 95, Applicatio	9.82e+01
17	56	34.1	1026	2	US-08-453-Sequence 95, Applicatio	9.82e+01
18	56	34.1	1026	3	PCT-US93-1Sequence 95, Applicatio	9.82e+01
19	56	34.1	1026	1	US-08-453-Sequence 95, Applicatio	9.82e+01
20	56	34.1	1203	1	US-08-453-Sequence 103, Applicat	9.82e+01
21	56	34.1	1203	1	US-08-268-Sequence 103, Applicat	9.82e+01
22	56	34.1	1203	1	US-08-453-Sequence 103, Applicat	9.82e+01
23	56	34.1	1203	3	PCT-US93-1Sequence 103, Applicat	9.82e+01

24	56	34.1	1203	1	US-07-998-Sequence 103, Applicat	9.82e+01
25	56	34.1	1203	3	PCT-US95-0Sequence 103, Applicat	9.82e+01
26	56	34.1	1203	2	US-08-453-Sequence 103, Applicat	9.82e+01
27	55	33.5	106	1	US-08-241-Sequence 12, Applicati	1.21e+02
28	55	33.5	106	2	US-08-850-Sequence 12, Applicati	1.21e+02
29	55	33.5	424	3	PCT-US94-1Sequence 18, Applicati	1.21e+02
30	55	33.5	3218	1	US-08-764-Sequence 27, Applicati	1.21e+02
31	54	32.9	189	2	US-08-637-Sequence 74, Applicati	1.49e+02
32	54	32.9	258	2	US-08-463-Sequence 10, Applicati	1.49e+02
33	54	32.9	258	2	US-08-462-Sequence 10, Applicati	1.49e+02
34	54	32.9	258	2	US-08-461-Sequence 10, Applicati	1.49e+02
35	54	32.9	556	2	US-08-505-Sequence 1, Applicatio	1.49e+02
36	54	32.9	560	2	US-08-981-Sequence 2, Applicatio	1.49e+02
37	54	32.9	1336	2	US-08-486-Sequence 58, Applicati	1.49e+02
38	54	32.9	1336	2	US-08-231-Sequence 52, Applicati	1.49e+02
39	53	32.3	56	2	US-08-537-Sequence 32, Applicati	1.83e+02
40	53	32.3	301	1	US-08-253-Sequence 43, Applicati	1.83e+02
41	53	32.3	390	2	US-08-979-Sequence 1, Applicatio	1.83e+02
42	53	32.3	908	3	PCT-US95-0Sequence 3, Applicatio	1.83e+02
43	52	31.7	98	2	US-08-485-Sequence 50, Applicati	2.25e+02
44	52	31.7	98	2	US-08-481-Sequence 50, Applicati	2.25e+02
45	52	31.7	568	2	US-08-782-Sequence 6, Applicatio	2.25e+02

ALIGNMENTS

RESULT 1
ID US-08-475-894-4 STANDARD; PRT; 464 AA.
AC xxxxxx
XX
XX
DT
XX
DE
XX
Sequence 4, Application US/08475894
Sequence 4, Application US/08475894
Patent No. 5641748
GENERAL INFORMATION:
APPLICANT: Yen-Ming Hsu
TITLE OF INVENTION: THE CAIP-LIKE GENE FAMILY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,894
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: BGP-191
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 464 amino acids
TYPE: amino acid
* TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
SEQUENCE 464 AA; 51267 MW; 1137110 CN;

[illegible]

CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/475,894
CC FILING DATE:
CC CLASSIFICATION: 530
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Louis Myers
CC REGISTRATION NUMBER: 35,965
CC REFERENCE/DOCKET NUMBER: BGP-191
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617)227-7400
CC TELEFAX: (617)227-5941
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 553 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 553 AA; 61925 MW; 1616798 CN;
SQ
Query Match 37.8%; Score 62; DB 1; Length 553;
Best Local Similarity 55.6%; Pred. No. 2.73e+01;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
Db 435 KIPPERPEMLPNTREEKE 452
QY 318 KTEAESPRMLDPQTEKE 335
RESULT 5
ID US-08-484-710-2 STANDARD; PRT; 553 AA.
XX
AC xxxxxx
XX
XX
DT
XX
DE
XX
Sequence 2, Application US/08484710
CC Sequence 2, Application US/08484710
CC Patent No. 5656438
CC GENERAL INFORMATION:
CC APPLICANT: Yen-Ming Hsu
CC TITLE OF INVENTION: THE CAIP-LIKE GENE FAMILY
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 State Street, Suite 510
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: USA
CC ZIP: 02109-1875
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/484,710
CC FILING DATE:
CC CLASSIFICATION: 530
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Louis Myers
CC REGISTRATION NUMBER: 35,965
CC REFERENCE/DOCKET NUMBER: BGP-190
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617)227-7400
CC TELEFAX: (617)227-5941
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 553 amino acids
CC TYPE: amino acid

CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 553 AA; 61925 MW; 1616798 CN;
SQ
Query Match 37.8%; Score 62; DB 1; Length 553;
Best Local Similarity 55.6%; Pred. No. 2.73e+01;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
Db 435 KIPPERPEMLPNTREEKE 452
QY 318 KTEAESPRMLDPQTEKE 335
RESULT 6
ID US-08-484-709-2 STANDARD; PRT; 553 AA.
XX
AC xxxxxx
XX
XX
DT
XX
DE
XX
Sequence 2, Application US/08484709
CC Sequence 2, Application US/08484709
CC Patent No. 5837844
CC GENERAL INFORMATION:
CC APPLICANT: Yen-Ming Hsu
CC TITLE OF INVENTION: THE CAIP-LIKE GENE FAMILY
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 State Street, Suite 510
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: USA
CC ZIP: 02109-1875
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/484,709
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Louis Myers
CC REGISTRATION NUMBER: 35,965
CC REFERENCE/DOCKET NUMBER: BGP-192
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617)227-7400
CC TELEFAX: (617)227-5941
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 553 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 553 AA; 61925 MW; 1616798 CN;
SQ
Query Match 37.8%; Score 62; DB 2; Length 553;
Best Local Similarity 55.6%; Pred. No. 2.73e+01;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
Db 435 KIPPERPEMLPNTREEKE 452
QY 318 KTEAESPRMLDPQTEKE 335
RESULT 7
ID US-08-769-309A-5 STANDARD; PRT; 1780 AA.
XX

```

AC      AC      xxxxxx
XX      DT      Sequence 5, Application US/08769309A
XX      DE      Sequence 5, Application US/08769309A
XX      CC      Patent No. 5741890
CC      CC      GENERAL INFORMATION:
CC      CC      APPLICANT: Scott, John D.,
CC      CC      APPLICANT: Nauert, Brian J.,
CC      CC      APPLICANT: Klauck, Theresa M.
CC      CC      TITLE OF INVENTION: Protein Binding Domains of Gravin
CC      CC      NUMBER OF SEQUENCES: 24
CC      CC      CORRESPONDENCE ADDRESS:
CC      CC      ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
CC      CC      STREET: 6300 Sears Tower/233 South Wacker Drive
CC      CC      CITY: Chicago
CC      CC      STATE: Illinois
CC      CC      COUNTRY: United States of America
CC      CC      ZIP: 60606-6402
CC      CC      COMPUTER READABLE FORM:
CC      CC      MEDIUM TYPE: Floppy disk
CC      CC      COMPUTER: IBM PC compatible
CC      CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      CC      SOFTWARE: PatentIn Release #1.0, Version #1.30
CC      CC      CURRENT APPLICATION DATA:
CC      CC      APPLICATION NUMBER: US/08/769,309A
CC      CC      FILING DATE:
CC      CC      CLASSIFICATION: 435
CC      CC      ATTORNEY/AGENT INFORMATION:
CC      CC      NAME: NO. 5741890and, Greta E.
CC      CC      REGISTRATION NUMBER: 35,302
CC      CC      REFERENCE/DOCKET NUMBER: 27866/33451
CC      CC      TELECOMMUNICATION INFORMATION:
CC      CC      TELEPHONE: 312-474-6300
CC      CC      TELEFAX: 312-474-0448
CC      CC      TELEX: 25-3856
CC      CC      INFORMATION FOR SEQ ID NO: 5:
CC      CC      SEQUENCE CHARACTERISTICS:
CC      CC      LENGTH: 1780 amino acids
CC      CC      TYPE: amino acid
CC      CC      TOPOLOGY: linear
CC      CC      MOLECULE TYPE: protein
SQ      SEQUENCE 1780 AA: 191552 MW; 16224320 CN;

Query Match          36.6% Score 60; DB 1: Length 1780;
Best Local Similarity 34.8%; Pred.No. 4.20e+01;
Matches              8; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Db    QY    942 AEEPTVTTEPLPENREAGDVT 964
       :| |: | |:| |:| |:|
QY    319 TEAESPRMLDPQTEEKESGGSL 341

RESULT      8
ID         US-08-4446-083-1. STANDARD; PRT; 189 AA.
XX
XX      xxxxxx
XX      DT
XX      DE
XX      CC
XX      CC      Sequence 1, Application US/08446083
XX      CC      Sequence 1, Application US/08446083
XX      CC      Patent No. 5804427
XX      CC      GENERAL INFORMATION:
XX      CC      APPLICANT: Davis, Roger J.
XX      CC      APPLICANT: Raineaud, Joel
XX      CC      APPLICANT: Gupta, Shashi
XX      CC      APPLICANT: Derlijard, Benoit
XX      CC      TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
XX      CC      TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE

```

CC APPLICATION NUMBER: US/08/933,821
CC FILING DATE:
CC CLASSIFICATION: 536
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Dreger, Ginger R.
CC REGISTRATION NUMBER: 33,055
CC REFERENCE/DOCKET NUMBER: P1130
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 650/225-3216
CC TELEFAX: 650/952-9881
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 470 amino acids
CC TYPE: Amino Acid
CC TOPOLOGY: Linear
CC SEQUENCE 470 AA; 51694 MW; 1069928 CN;

Query Match 36.0%; Score 59; DB 2; Length 470;
Best Local Similarity 33.3%; Pred. No. 5.20e+01;
Matches 6; Conservative 8; Mismatches 3; Indels 1; Gaps 1;

Db 214 SDTSRMLDPAPERDOT 231
QY 321 AESPRMLDPQTE-EKEAS 337

RESULT 10
ID PCT-US94-00198-3 STANDARD; PRT; 2938 AA.
XX AC xxxxxx
XX DT
XX DE
XX SEQUENCE 3, Application PC/TUS9400198
XX
XX SEQUENCE 3, Application PC/TUS9400198
XX GENERAL INFORMATION:
XX APPLICANT: Schering Corp.
XX TITLE OF INVENTION: RAS Associated GAP Proteins
XX NUMBER OF SEQUENCES: 6
XX CORRESPONDENCE ADDRESS:
XX ADDRESSEE: Schering Corp.
XX STREET: 1 Giraldo Farms
XX CITY: Madison
XX STATE: New Jersey
XX COUNTRY: USA
XX ZIP: 94304-1104
XX COMPUTER READABLE FORM:
XX MEDIUM TYPE: Floppy disk
XX COMPUTER: Macintosh
XX OPERATING SYSTEM: 6.0.8
XX SOFTWARE: Microsoft Word 5.1a
XX CURRENT APPLICATION DATA:
XX APPLICATION NUMBER: PCT/US94/00198
XX FILING DATE:
XX CLASSIFICATION:
XX PRIOR APPLICATION DATA:
XX APPLICATION NUMBER: 3:
XX APPLICATION NUMBER: 824
XX FILING DATE: 15-JAN-1993
XX ATTORNEY/AGENT INFORMATION:
XX NAME: Lunn, Paul G.
XX REGISTRATION NUMBER: 32,743
XX REFERENCE/DOCKET NUMBER: DX0352 PCT
XX TELEPHONE: (201)822-7255
XX TELEFAX: (201)822-7039
XX INFORMATION FOR SEQ ID NO: 3:
XX SEQUENCE CHARACTERISTICS:
XX LENGTH: 2938 amino acids
XX TYPE: amino acid
XX STRANDEDNESS: single
XX TOPOLOGY: linear
XX MOLECULE TYPE: protein

CC ORIGINAL SOURCE:
CC ORGANISM: Saccharomyces cerevisiae
CC SEQUENCE 2938 AA; 333115 MW; 46612862 CN;

Query Match 36.0%; Score 59; DB 3; Length 2938;
Best Local Similarity 20.0%; Pred. No. 5.20e+01;
Matches 5; Conservative 14; Mismatches 6; Indels 0; Gaps 0;

Db 310 TKTDADTPSTMTNNNNNNNSANL 334
QY 317 AKTEASPRMLDPQTEKEASGSL 341

RESULT 11
ID US-08-805-965-3 STANDARD; PRT; 268 AA.
XX AC xxxxxx
XX DT
XX DE
XX SEQUENCE 3, Application US/08805965
XX
XX SEQUENCE 3, Application US/08805965
XX Patent No. 5851774
XX GENERAL INFORMATION:
XX APPLICANT: Hillman, Jennifer L.
XX APPLICANT: Goli, Surya K.
XX TITLE OF INVENTION: NOVEL HUMAN MLS3 PROTEIN
XX NUMBER OF SEQUENCES: 6
XX CORRESPONDENCE ADDRESS:
XX ADDRESSEE: Incyte Pharmaceuticals, Inc.
XX STREET: 3174 Porter Drive
XX CITY: Palo Alto
XX STATE: CA
XX COUNTRY: USA
XX ZIP: 94304
XX COMPUTER READABLE FORM:
XX MEDIUM TYPE: Diskette
XX COMPUTER: IBM Compatible
XX OPERATING SYSTEM: DOS
XX SOFTWARE: FastSeq for Windows Version 2.0
XX CURRENT APPLICATION DATA:
XX APPLICATION NUMBER: US/08/805,965
XX FILING DATE: Herewith
XX CLASSIFICATION: 424
XX PRIOR APPLICATION DATA:
XX APPLICATION NUMBER:
XX FILING DATE:
XX ATTORNEY/AGENT INFORMATION:
XX NAME: Billings, Lucy J.
XX REGISTRATION NUMBER: 36,749
XX REFERENCE/DOCKET NUMBER: PF-0223 US
XX TELECOMMUNICATION INFORMATION:
XX TELEPHONE: 415-855-0555
XX TELEFAX: 415-845-4166
XX TELEX:
XX INFORMATION FOR SEQ ID NO: 3:
XX SEQUENCE CHARACTERISTICS:
XX LENGTH: 268 amino acids
XX TYPE: amino acid
XX STRANDEDNESS: single
XX TOPOLOGY: linear
XX IMMEDIATE SOURCE:
XX LIBRARY: GenBank
XX CLONE: 1066392
XX SEQUENCE 268 AA; 30627 MW; 349374 CN;

Query Match 34.1%; Score 56; DB 2; Length 268;
Best Local Similarity 30.4%; Pred. No. 9.82e+01;
Matches 7; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

Db 114 SKIGDEPPKVFQASTOTRRAPGG 136
:| :|:|:|:| :|:|:|

QY 317 AKTEASPRMLDPQTEKEASGG 339

RESULT 12
ID US-09-057-762-7 STANDARD; PRT; 335 AA.
XX
AC xxxxxx
DT
XX
DE
XX
XX
XX
XX
XX
Sequence 7, Application US/09057762
CC
XX
Sequence 7, Application US/09057762
CC
Patent No. 5879909
CC
GENERAL INFORMATION:
CC
APPLICANT: PERL, ANDRAS
CC
TITLE OF INVENTION: HUMAN TRANSALDOLASE: AN AUTOANTIGEN WITH
CC
TITLE OF INVENTION: A FUNCTION IN METABOLISM
CC
NUMBER OF SEQUENCES: 24
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: MORRISON & FOERSTER
CC
STREET: 2000 Pennsylvania Avenue N.W.
CC
CITY: Washington
CC
STATE: D.C.
CC
COUNTRY: USA
CC
ZIP: 20006-1812
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Floppy disk
CC
COMPUTER: IBM PC compatible
CC
OPERATING SYSTEM: PC-DOS/MS-DOS
CC
SOFTWARE: Patent in Release #1.0, Version #1.30
CC
CURRENT APPLICATION DATA:
CC
APPLICATION NUMBER: US/09/057.762
CC
FILING DATE:
CC
CLASSIFICATION: 530
CC
PRIOR APPLICATION NUMBER: US 08/326.119
CC
FILING DATE: 19-OCT-1994
CC
ATTORNEY/AGENT INFORMATION:
CC
NAME: LIVNAT, SHMUEL
CC
REGISTRATION NUMBER: 33,949
CC
REFERENCE/DOCKET NUMBER: 280932000100
CC
TELECOMMUNICATION INFORMATION:
CC
TELEPHONE: (202) 887-1500
CC
TELEFAX: (202) 822-0168
CC
TELEX: 90-4030
CC
INFORMATION FOR SEQ ID NO: 7:
CC
SEQUENCE CHARACTERISTICS:
CC
LENGTH: 335 amino acids
CC
TYPE: amino acid
CC
STRANDEDNESS: single
CC
TOPOLOGY: linear
CC
SEQUENCE 335 AA; 37054 MW; 596293 CN;

Query Match 34.1%; Score 56; DB 2; Length 335;
Best Local Similarity 57.1%; Pred. No. 9.82e+01;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 272 PRVLDPVSAKKEAG 285
QY 324 PRMLDPQTEKEAS 337

RESULT 13
ID US-08-268-161A-95 STANDARD; PRT; 1026 AA.
XX
AC xxxxxx
DT
XX
DE
XX
XX
Sequence 95, Application US/08268161A
CC
XX
Sequence 95, Application US/08268161A

Patent No. 5798224
GENERAL INFORMATION:
CC
APPLICANT: Suzuki, Shintaro
CC
TITLE OF INVENTION: Protocadherin Materials and Methods
CC
NUMBER OF SEQUENCES: 115
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
CC
STREET: 233 South Wacker, 6300 Sears Tower
CC
CITY: Chicago
CC
STATE: Illinois
CC
COUNTRY: USA
CC
ZIP: 60606
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Floppy disk
CC
COMPUTER: IBM PC compatible
CC
OPERATING SYSTEM: PC-DOS/MS-DOS
CC
SOFTWARE: Patent in Release #1.0, Version #1.25
CC
CURRENT APPLICATION DATA:
CC
APPLICATION NUMBER: US/08/268,161A
CC
FILING DATE: June 27, 1994
CC
CLASSIFICATION: 435
CC
ATTORNEY/AGENT INFORMATION:
CC
NAME: Young J. Suh
CC
REGISTRATION NUMBER: P-41,337
CC
REFERENCE/DOCKET NUMBER: 27866/32149
CC
TELECOMMUNICATION INFORMATION:
CC
TELEPHONE: 312/474-6300
CC
TELEFAX: 312/474-0448
CC
TELEX: 25-3856
CC
INFORMATION FOR SEQ ID NO: 95:
CC
SEQUENCE CHARACTERISTICS:
CC
LENGTH: 1026 amino acids
CC
TYPE: amino acid
CC
TOPOLOGY: linear
CC
MOLECULE TYPE: protein
CC
SEQUENCE 1026 AA; 111270 MW; 5611711 CN;

Query Match 34.1%; Score 56; DB 1; Length 1026;
Best Local Similarity 28.6%; Pred. No. 9.82e+01;
Matches 6; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Db 123 QNASPRLLGQIEVDINDNT 143
QY 320 EABSPRMLDPQTEKEASGG 340

RESULT 14
ID US-08-453-274B-95 STANDARD; PRT; 1026 AA.
XX
AC xxxxxx
DT
XX
DE
XX
XX
Sequence 95, Application US/08453274B
CC
Patent No. 5663300
CC
GENERAL INFORMATION:
CC
APPLICANT: Suzuki, Shintaro
CC
TITLE OF INVENTION: Protocadherin Materials and Methods
CC
NUMBER OF SEQUENCES: 107
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
CC
STREET: 6300 Sears Tower, 233 South Wacker Drive
CC
CITY: Chicago
CC
STATE: Illinois
CC
COUNTRY: United States of America
CC
ZIP: 60606-6402
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Floppy disk
CC
COMPUTER: IBM PC compatible
CC
OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: PatentIn Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA: US/08/453,274B
 CC FILING DATE: 30-MAY-1995
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: No. 5663300and, Greta E.
 CC REGISTRATION NUMBER: 35,302
 CC REFERENCE/DOCKET NUMBER: 32660
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 312/474-6300
 CC TELEFAX: 312/474-0448
 CC TELEX: 25-3856
 CC INFORMATION FOR SEQ ID NO: 95:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 1026 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC SEQUENCE 1026 AA; 111270 MW; 5611711 CN;
 CC
 CC Query Match 34.1%; Score 56; DB 1; Length 1026;
 CC Best Local Similarity 28.6%; Pred. No. 9.82e+01;
 CC Matches 6; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
 CC
 Db 123 QNASPRLLGGQIEVODINDNT 143
 QY 320 EAESPRMLDPQTEKEASGGS 340
 CC
 RESULT 15
 ID US-07-998-003A-95 STANDARD; PRT; 1026 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX
 XX
 DE Sequence 95, Application US/07998003A
 CC Sequence 95, Application US/07998003A
 CC Patent No. 5643781
 CC GENERAL INFORMATION:
 CC APPLICANT: Suzuki, Shintaro
 CC TITLE OF INVENTION: Protocadherin Materials and Methods
 CC NUMBER OF SEQUENCES: 107
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
 CC ADDRESSEE: Bicknell
 CC STREET: 20 South Clark Street
 CC CITY: Chicago
 CC STATE: Illinois
 CC COUNTRY: USA
 CC ZIP: 60603
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/07/998,003A
 CC FILING DATE:
 CC CLASSIFICATION: 435
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: No. 564378land, Greta E.
 CC REGISTRATION NUMBER: 35,302
 CC REFERENCE/DOCKET NUMBER: 30903
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 312/346-5750
 CC TELEFAX: 312/984-9740
 CC TELEX: 25-3856
 CC INFORMATION FOR SEQ ID NO: 95:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 1026 amino acids
 CC TYPE: amino acid

CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC SEQUENCE 1026 AA; 111270 MW; 5611711 CN;
 CC
 CC Query Match 34.1%; Score 56; DB 1; Length 1026;
 CC Best Local Similarity 28.6%; Pred. No. 9.82e+01;
 CC Matches 6; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
 CC
 Db 123 QNASPRLLGGQIEVODINDNT 143
 QY 320 EAESPRMLDPQTEKEASGGS 340
 CC
 Search completed: Wed May 10 14:15:16 2000
 Job time : 57 secs.

W P E R E H (TW)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, G.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 14:13:27 2000; MasPar time 5.02 seconds
Tabular output not generated. 244.109 Million cell updates/sec

Title: >US-09-376-430-2
Description: (317-342) from US09376430A.pep (23 of 25)
Perfect Score: 164
Sequence: 1 AKTEASPRMLDPQTEKEASGSLQ 26

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 28.807; Variance 43.668; scale 0.660

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	66	40.2	367	2 H70519	probable aceAa protei	1.44e+00
2	66	40.2	606	2 S77654	isocitrate lyase (EC	1.44e+00
3	63	38.4	362	2 G9589	3-dehydroquinate synt	4.66e+00
4	63	38.4	379	2 S42543	hypothetical protein	4.66e+00
5	63	38.4	506	2 F70617	hypothetical protein	4.66e+00
6	63	38.4	828	1 D39142	outer membrane usher	4.66e+00
7	62	37.8	959	2 S48962	MSH1 protein - yeast	6.83e+00
8	62	37.8	1616	2 G54242	cytadherence-accessor	6.83e+00
9	62	37.8	2346	2 T13829	tpi homolog - fruit f	6.83e+00
10	61	37.2	371	2 T06382	knox protein 1 - gard	9.98e+00
11	61	37.2	390	1 A42724	transcription initiat	9.98e+00
12	61	37.2	811	2 A41054	fasciclin II, transme	9.98e+00
13	61	37.2	873	2 B41054	fasciclin II PI-linke	9.98e+00
14	61	37.2	882	2 S77704	hypothetical protein	9.98e+00
15	60	36.6	294	2 A41061	probable transcriptio	1.45e+01
16	60	36.6	294	2 S32234	transcription antiter	1.45e+01
17	60	36.6	300	2 A41062	probable transcriptio	1.45e+01
18	60	36.6	300	2 S54717	probable transcriptio	1.45e+01
19	60	36.6	321	2 S51395	hypothetical protein	1.45e+01
20	60	36.6	404	2 A28404	S-antigen - bovine	1.45e+01
21	60	36.6	455	2 S24458	hypothetical protein	1.45e+01
22	60	36.6	1239	1 A32579	neuroglian - fruit fl	1.45e+01
23	60	36.6	1684	2 JW0057	gravin - human	1.45e+01

24	60	36.6	2489	2 S59782	probable membrane pro	1.45e+01
25	59	36.0	178	2# S54444	prelin 18.1K type 4	2.10e+01
26	59	36.0	299	2 T11788	probable transcriptio	2.10e+01
27	59	36.0	319	2 A41773	butyrolactone autoreg	2.10e+01
28	59	36.0	739	1 VHIWEB	nucleocapsid protein	2.10e+01
29	59	36.0	2137	1 SJHUB	spectrin beta chain -	2.10e+01
30	59	36.0	3092	2 S46009	GTPase-activating pro	2.10e+01
31	58	35.4	328	2 A46521	52K phosphoprotein -	3.02e+01
32	58	35.4	330	2 A30533	lymphocyte-specific p	3.02e+01
33	58	35.4	330	2 I57835	lymphocyte-specific p	3.02e+01
34	58	35.4	348	1 R5HS10	ribosomal protein L10	3.02e+01
35	58	35.4	674	1 A41670	carbon-monoxide denyd	3.02e+01
36	58	35.4	714	2 S68603	hypothetical protein	3.02e+01
37	58	35.4	817	2 JC4176	pyruvate water dikina	3.02e+01
38	58	35.4	982	2 S09810	hypothetical protein	3.02e+01
39	58	35.4	1181	2 T01799	hypothetical protein	3.02e+01
40	58	35.4	1234	2 G70622	probable transcriptio	3.02e+01
41	58	35.4	1584	2 T00026	brain-specific angio	3.02e+01
42	58	35.4	1887	2 S61703	fatty-acid synthase (3.02e+01
43	57	34.8	175	2 S48795	troponin T, cardiac m	4.33e+01
44	57	34.8	203	2 E75145	lsu ribosomal protein	4.33e+01
45	57	34.8	912	2 D72644	hypothetical protein	4.33e+01

ALIGNMENTS

RESULT 1	H70519	#type complete	
ENTRY	probable aceAa protein - Mycobacterium tuberculosis (strain H37RV)		
TITLE	#formal_name Mycobacterium tuberculosis		
ORGANISM	17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change		
DATE	17-Jul-1998		
ACCESSIONS	H70519		
REFERENCE	A70500		
#authors	Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.		
#journal	Nature (1998) 393:537-544		
#title	Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.		
#cross-references	MUID:98295987		
#accession	H70519		
##status	preliminary; nucleic acid sequence not shown;		
##molecule_type	DNA		
##residues	1-367 ##label COL		
##cross-references	GB:297193; GB:AL123456; NID:g3261816; PID:e324889; PID:g2225945		
##experimental_source	strain H37RV		
GENETICS			
#gene	aceAa		
SUMMARY	#length 367 #molecular-weight 40489 #checksum 4308		
Query Match	40.2%; Score 66; DB 2; Length 367;		
Best Local Similarity	31.8%; Pred.No. 1.44e+00;		
Matches	6; Conservative 12; Mismatches 1; Indels 0; Gaps 0;		
Db	251 ARTDAPANLIDSRABERD 269		
QV	317 AKWEASPRMLDPQTEKE 335		
RESULT 2	S77654	#type complete	
ENTRY	isocitrate lyase (EC 4.1.3.1) - Mycobacterium leprae		
TITLE			

```

Terpstra, P.; Toggoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession G69589
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-362 ##label KUN
##cross-references GB:299115; NID:AL009126; PID:e1183715;
PID:G2634688
##experimental_source strain 168
GENETICS
#gene aroB
CLASSIFICATION
#superfamily 3-dehydroquinase synthase; 3-dehydroquinase
synthase homology
FEATURE
5-337
SUMMARY
#domain 3-dehydroquinase synthase homology #label DQS
#length 362 #molecular-weight 40817 #checksum 5049
Query Match 38.4%; Score 63; DB 2; Length 362;
Best Local Similarity 28.0%; Pred.No. 4.66e+00;
Matches 7; Conservative 10; Mismatches 8; Indels 0; Gaps 0;
Db 305 RKETSTSVLLNRMMNDKRTGGKIQ 329
:|:::|:::|:::|:::|:::|
QY 318 KTEASPRMLDPQTEKEASGGS LQ 342
RESULT 4
ENTRY S42543 #type complete
TITLE hypothetical protein - soybean
ORGANISM #formal_name Glycine max #common_name soybean
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
24-Sep-1999
ACCESSIONS S42543
REFERENCE S42543
#authors Ma, H.; McMullen, M.D.; Finer, J.J.
#journal Plant Mol. Biol. (1994) 24:465-473
#title Identification of a homeobox-containing gene with enhanced
expression during soybean (Glycine max L.) somatic embryo
development.
#cross-references MUID:94169300
#accession S42543
##status preliminary
##molecule_type mRNA
##residues 1-379 ##label MAH
##cross-references EMBL:L13663; NID:g310568; PIDN:AAA20882.1;
PID:g485406
CLASSIFICATION #superfamily unassigned homeobox proteins; homeobox homology
KEYWORDS DNA binding; homeobox; nucleus; transcription regulation
FEATURE
284-343
SUMMARY
#domain homeobox homology #label HOX
#length 379 #molecular-weight 42374 #checksum 7237
Query-Match 38.4%; Score 63; DB 2; Length 379;
Best Local Similarity 30.4%; Pred.No. 4.66e+00;
Matches 7; Conservative 8; Mismatches 8; Indels 0; Gaps 0;
Db 246 DVDLHNMDPQAEEDRLKGLLR 268
: : |::|::|:::|:::|:::|
QY 320 EAESPRMLDPQTEKEASGGS LQ 342
RESULT 5
ENTRY F70617 #type complete
TITLE hypothetical protein Rv0147 - Mycobacterium tuberculosis

```

```

ORGANISM      (strain H37RV)
DATE          #formal_name Mycobacterium tuberculosis
              17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
              20-Sep-1999
ACCESSIONS    F70617
REFERENCE      A70500
#authors      Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
              C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
              III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
              Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
              Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
              Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
              Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
              Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
              Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
              Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal      Nature (1998) 393:537-544
#title        Deciphering the biology of Mycobacterium tuberculosis from
              the complete genome sequence.
#cross-references MUID:98295987
#accession    F70617
#status       preliminary; nucleic acid sequence not shown;
              translation not shown
#molecule_type DNA
#residues     1-506 ##label COL
#cross-references GB:292770; GB:AL123456; NID:g3261720; PID:e306525;
              PID:g1877272
#experimental_source strain H37RV
GENETICS      RV0147
#gene         #superfamily aldehyde dehydrogenase (NAD+); aldehyde
              dehydrogenase homology
CLASSIFICATION #length 506 #molecule_weight 55035 #checksum 2270
SUMMARY
Query Match   38.4%; Score 63; DB 2; Length 506;
Best Local Similarity 42.1%; Pred. No. 4.66e+00;
Matches 8; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
Db 28 KSETEAPREPAPVSEKOT 46
|:|:| | | | | |
Qy 318 KTEAESPRMLDPQTEKEA 336
|:|:| | | | | |
RESULT        6
ENTRY         #type complete
TITLE         outer membrane usher protein mrkC precursor - Klebsiella
              pneumoniae
ORGANISM      #formal_name Klebsiella pneumoniae
DATE          10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
ACCESSIONS    D39142
REFERENCE      A39142
#authors      Allen, B.L.; Gerlach, G.F.; Clegg, S.
#journal      J. Bacteriol. (1991) 173:916-920
#title        Nucleotide sequence and functions of mrk determinants
              necessary for expression of type 3 fimbriae in Klebsiella
              pneumoniae.
#cross-references MUID:91100388
#accession    D39142
#status       preliminary
#molecule_type DNA
#residues     1-828 ##label ALL
#cross-references GB:M55912; NID:g149234; PIDN:AAA25095.1; PID:g149238
CLASSIFICATION #superfamily outer membrane usher protein fimb
KEYWORDS      membrane protein
#length 828 #molecule_weight 91049 #checksum 429
SUMMARY
Query Match   38.4%; Score 63; DB 1; Length 828;
Best Local Similarity 55.0%; Pred. No. 4.66e+00;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
Db 106 SPEKLNPNQLSEKGEFGRLQ 125
|:|:| | | | | |

```

```

Qy 323 SPRMLDPQTEKEASGGSLQ 342
|:|:| | | | | |
RESULT        7
ENTRY         #type complete
TITLE         MSH1 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein YHR120w
ORGANISM      #formal_name Saccharomyces cerevisiae
DATE          28-Oct-1994 #sequence_revision 27-Jan-1995 #text_change
              07-May-1999
ACCESSIONS    S48962; S50781; S27432
REFERENCE      S46591
#authors      Latreille, P.
#submission   submitted to the EMBL Data Library, May 1994
#description   The sequence of S. cerevisiae cosmid 8263.
#accession    S48962
#molecule_type DNA
#residues     1-959 ##label LAT
#cross-references EMBL:U00059; NID:g529116; PID:g529134; MIPS:YHR120w
REFERENCE      S50781
#authors      Reenan, R.A.G.; Kolodner, R.D.
#journal      Genetics (1992) 132:963-973
#title        Isolation and characterization of two Saccharomyces
              cerevisiae genes encoding homologs of the bacterial hexa
              and muts mismatch repair proteins.
#cross-references MUID:93093456
#accession    S50781
#molecule_type DNA
#residues     1-436, 'N', 438-694, 'V', 696-959 ##label REE
#cross-references EMBL:M84169; NID:g171999; PID:g172000
GENETICS      SGD:MSH1
#gene         #cross-references SGD:S0001162; MIPS:YHR120w
#map_position 9R
KEYWORDS      DNA binding
SUMMARY       #length 959 #molecular_weight 109407 #checksum 1491
Query Match   37.8%; Score 62; DB 2; Length 959;
Best Local Similarity 42.9%; Pred. No. 6.83e+00;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
Db 529 AKQNAVTRMLDIDVKKES 549
|:|:| | | | | |
Qy 317 AKTEAESPRMLDPQTEKEAS 337
|:|:| | | | | |
RESULT        8
ENTRY         #type complete
TITLE         cytoadherence-accessory protein (hmw1) homolog MG386 -
              Mycoplasma genitalium (SGC3)
ORGANISM      #formal_name Mycoplasma genitalium
DATE          17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change
              10-Oct-1997
ACCESSIONS    G64242
REFERENCE      A64200
#authors      Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton,
              R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.;
              Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.;
              Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.;
              Uutterback, I.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
              J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.;
              Luciet, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III,
              C.A.; Venter, J.C.
#journal      Science (1995) 270:397-403
#title        The minimal gene complement of Mycoplasma genitalium.
#cross-references MUID:96026346
#accession    G64242
#status       preliminary; nucleic acid sequence not shown;
              translation not shown
#molecule_type DNA
#residues     1-1616 ##label TIGR
#cross-references GB:U39723; GB:L43967; NID:g1046092; PID:g1046097;
              TIGR:MG386

```

##experimental_source strain G-37

GENETICS

#genetic_code SGC3

SUMMARY #length 1616 #molecular-weight 185677 #checksum 3614

Query Match 37.8%; Score 62; DB 2; Length 1616;
Best Local Similarity 42.9%; Pred. No. 6.83e+00;
Matches 9; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 803 ESNALFDSKPKDEKSSDSEIQ 823

|| :||: ||||: ||

QY 322 ESPRMLDPQTEKEASGSLQ 342

RESULT 9

ENTRY #type complete
TITLE Tpr homolog - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

ACCESSIONS

REFERENCE T13829

#authors Zimowska, G.; Aris, J.P.; Paddy, M.R.

#journal J. Cell Sci. (1997) 110:927-944

#title A Drosophila Tpr protein homolog is localized both in the extrachromosomal channel network and to nuclear pore complexes.

#accession

T13829

##status preliminary; translated from GB/EMBL/DBJ

##molecule_type mRNA

##residues 1-2346 #label ZIM

##cross-references EMBL:U91980; NID:gl923273; PID:gl923274; PIDN:AAC47506.1

GENETICS

#map_position 2R

SUMMARY #length 2346 #molecular-weight 262363 #checksum 5206

Query Match 37.8%; Score 62; DB 2; Length 2346;
Best Local Similarity 32.0%; Pred. No. 6.83e+00;
Matches 8; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

Db 1121 RSNAAQKLLDKRESEKRVSLH 1145

||||: ||||: ||||: ||

QY 318 KTEASPRMLDPQTEKEASGSLQ 342

RESULT 10

ENTRY #type complete
TITLE Knox protein 1 - garden pea
ORGANISM #formal_name Pisum sativum #common_name garden pea
DATE 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 24-Sep-1999

ACCESSIONS

REFERENCE T06382

#authors Giles, J.E.; Villani, P.J.; DeMason, D.A.

#submission Submitted to the EMBL Data Library, May 1998

#accession T06382

##status preliminary; translated from GB/EMBL/DBJ

##molecule_type mRNA

##residues 1-371 #label GIL

##cross-references EMBL:AF063307; NID:g3426303; PIDN:AAC32262.1; PID:g3426304

GENETICS

#gene Hop1

CLASSIFICATION #superfamily unassigned homeobox proteins; homeobox homology

SUMMARY #length 371 #molecular-weight 41854 #checksum 962

Query Match 37.2%; Score 61; DB 2; Length 371;
Best Local Similarity 47.1%; Pred. No. 9.98e+00;
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 244 MIDPOAEDRELKQLLR 260

||||: ||||: ||||: ||

QY 326 MLDPQTEKEASGSLQ 342

RESULT 11

ENTRY #type complete
TITLE transcription initiation factor sigma sign - Anabaena sp.
ORGANISM #formal_name Anabaena sp.
DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

ACCESSIONS

REFERENCE A42724

#authors Brahamsha, B.; Haselkorn, R.

#journal J. Bacteriol. (1991) 173:2442-2450

#title Isolation and characterization of the gene encoding the principal sigma factor of the vegetative cell RNA polymerase from the cyanobacterium Anabaena sp. strain PCC 7120.

##cross-references M01D:91193199

#accession A42724

##status preliminary

##molecule_type DNA

##residues 1-390 #label BRA

##cross-references GB:M60046; NID:gl42107; PIDN:AAA22043.1; PID:gl42108

CLASSIFICATION #superfamily transcription initiation factor sigma 43;

transcription initiation factor sigma katF homology;

transcription initiation factor sigma region 1 homology

KEYWORDS DNA binding; sigma factor; transcription initiation

FEATURE 160-385 #domain transcription initiation factor sigma katF

SUMMARY #length 390 #molecular-weight 45641 #checksum 8953

Query Match 37.2%; Score 61; DB 1; Length 390;
Best Local Similarity 35.0%; Pred. No. 9.98e+00;
Matches 7; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Db 42 DGDIDEFLEPOTDEDDAKSG 61

||: ||||: ||||: ||

QY 320 EAESPRMLDPQTEKEASGG 339

RESULT 12

ENTRY #type complete
TITLE fasciclin II, transmembrane splice form precursor - fruit fly
ORGANISM (Drosophila melanogaster)
DATE 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change 24-Sep-1998

ACCESSIONS

REFERENCE A41054

#authors Greeningloh, G.; Rehm, E.J.; Goodman, C.S.

#journal Cell (1991) 67:45-57

#title Genetic analysis of growth cone guidance in Drosophila:

fasciclin II functions as a neuronal recognition molecule.

##cross-references M01D:92005695

#accession A41054

##status preliminary

##molecule_type mRNA

##residues 1-811 #label GRE

##cross-references GB:M77165; NID:gl57402; PID:gl57403

GENETICS

#gene FlyBase:Fas2

##cross-references FlyBase:FBgn0000635

membrane protein

SUMMARY #length 811 #molecular-weight 90067 #checksum 1695

Query Match 37.2%; Score 61; DB 2; Length 811;
Best Local Similarity 36.0%; Pred. No. 9.98e+00;
Matches 9; Conservative 7; Mismatches 7; Indels 2; Gaps 2;

Db 369 QDDDPRIILEPFDERGESTGLR 393

: ||: ||||: ||||: ||

```

QY 320 EAESPRM-LDPO-TEEKEASGSLQ 342

RESULT 13
ENTRY   B41054      #type complete
TITLE   fasciclin II Pr-linked splice form precursor - fruit fly
        (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE     21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change
        16-Feb-1997

ACCESSIONS B41054
REFERENCE  A41054
#authors  Grenningloh, G.; Rehm, E.J.; Goodman, C.S.
#journal  Cell (1991) 67:45-57
#title    Genetic analysis of growth cone guidance in Drosophila:
#          fasciclin II functions as a neuronal recognition molecule.
#cross-references MUID:92005695
#accession B41054
#status     preliminary
#molecule_type mRNA
#residues   1-873 #label GRE
#cross-references GB:M77166

GENETICS
#gene       FlyBase:Fas2
#cross-references FlyBase:FBgn0000635
KEYWORDS    transmembrane protein
SUMMARY     #length 873 #molecular-weight 96911 #checksum 9845

Query Match      37.2%; Score 61; DB 2; Length 873;
Best Local Similarity 36.0%; Pred. No. 9.98e+00;
Matches 9; Conservative 7; Mismatches 7; Indels 2; Gaps 2;

Db 369 QDDPRILFNFDERGESTGTLR 393
: : ||: ||: ||: ||: ||:
QY 320 EAESPRM-LDPO-TEEKEASGSLQ 342

RESULT 14
ENTRY   S57704      #type complete
TITLE   hypothetical protein YGR238c - yeast (Saccharomyces
        cerevisiae)
ALTERNATE_NAMES hypothetical protein G8585
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     19-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
        24-Sep-1999
ACCESSIONS S57704; S64562; S63920
REFERENCE  S57680
#authors  van der Aart, Q.J.M.; Kleine, K.; Steensma, H.Y.
#submission submitted to the EMBL Data Library, June 1995
#description Sequence analysis of the 43 KB CRM1-YLM9-PET54-
        SM11-PHO81-YHB4-PFK1 region from the right arm of
        Saccharomyces cerevisiae chromosome VII.
#accession S57704
#molecule_type DNA
#residues   1-882 #label VAN
#cross-references EMBL:X87941; NID:g886908; PIDN:CAA61189.1;
        PID:g886933
#experimental_source strain S288C

REFERENCE S64541
#authors  van der Aart, Q.J.M.; Steensma, H.Y.
#submission submitted to the Protein Sequence Database, May 1996
#accession S64562
#molecule_type DNA
#residues   1-882 #label VAW
#cross-references EMBL:273023; NID:g1323430; PIDN:CAA97266.1;
        PID:e243674; PID:g1323431; MIPS:YGR238C
#experimental_source strain S288C

REFERENCE S63896
#authors  van der Aart, Q.J.M.; Kleine, K.; Steensma, H.Y.
#journal  Yeast (1996) 12:385-390
#title    Sequence analysis of the 43 kb
        CRM1-YLM9-PET54-DIF2-SM11-PHO81-YHB4-PFK1 region from the
        right arm of Saccharomyces cerevisiae chromosome VII.

```

```

#cross-references MUID:96267763
#accession S63920
#status     nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues   1-882 #label VAF
#cross-references EMBL:X87941; NID:g886908; PIDN:CAA61189.1;
        PID:g886933
#note       the nucleotide sequence was submitted to the EMBL Data
        Library, June 1995

GENETICS
#gene       SGD:KEL2
#cross-references SGD:S0003470; MIPS:YGR238C
#map_position 7R
CLASSIFICATION #superfamily hypothetical protein YHR158c
SUMMARY        #length 882 #molecular-weight 99974 #checksum 6435

Query Match      37.2%; Score 61; DB 2; Length 882;
Best Local Similarity 41.7%; Pred. No. 9.98e+00;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Db 379 KTDYASPNIHDLQSTEDQEGTL 402
||: ||: ||: ||: ||: ||:
QY 318 KTEASPRMLDPQTEKEASGSL 341

RESULT 15
ENTRY   S41061      #type complete
TITLE   Probable transcription antitermination factor nusG -
        Streptomyces griseus (strain IF013350)
ORGANISM #formal_name Streptomyces griseus
        strain IF013350
#variety  19-Mar-1997 #sequence_revision 12-Dec-1997 #text_change
        26-Aug-1999
ACCESSIONS S41061
REFERENCE  S41059
#authors  Miyake, K.; Onaka, H.; Horinouchi, S.; Beppu, T.
#journal  Biochim. Biophys. Acta (1994) 1217:97-100
#title    Organization and nucleotide sequence of the secE-nusG region
        of Streptomyces griseus.
#cross-references MUID:94114580
#accession S41061
#molecule_type DNA
#residues   1-294 #label MY
#cross-references EMBL:D17464; NID:g436786; PIDN:BAA04281.1;
        PID:g1004801; PID:g483836
#experimental_source strain IF013350

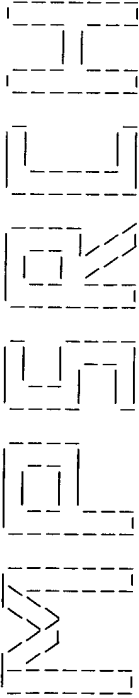
GENETICS
#gene       nusG
#start_codon GTG
FUNCTION     may be involved in antibiotics production
#description #superfamily transcription antitermination factor nusG
CLASSIFICATION #length 294 #molecular-weight 31939 #checksum 3281
SUMMARY

Query Match      36.6%; Score 60; DB 2; Length 294;
Best Local Similarity 43.8%; Pred. No. 1.45e+01;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 207 KMLAPEAEKKAAREAA 222
: || ||: ||: ||: ||: ||:
QY 325 RMLDPQTEKEASGGS 340

Search completed: Wed May 10 14:13:36 2000
Job time : 9 secs.

```



 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed May 10 14:07:12 2000; MasPar time 92.54 seconds
 Tabular output not generated. 8.557 Million cell updates/sec

Title: >US-09-376-430-2
 Description: (317-342) from US09376430A.pap (23 of 25)
 Perfect Score: 164
 Sequence: 1 AKTEASPRMLDPQTEKEASGGSLQ 26

Scoring table: PAM 150
 Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot38
 1:swissprot

Statistics: Mean 29.528; Variance 40.897; scale 0.722

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	67	40.9	1005	1	GCP_CAEEL	3.68e-01
2	66	40.2	606	1	ACEA_MYCLE	5.63e-01
3	63	38.4	362	1	AROB_BACSU	1.97e+00
4	63	38.4	379	1	HMB1_SOYBN	1.97e+00
5	63	38.4	828	1	MRKC_KLEPN	1.97e+00
6	63	38.4	1411	1	TCOF_HUMAN	1.97e+00
7	62	37.8	484	1	PAP2_XENLA	2.97e+00
8	62	37.8	959	1	MSH1_YEAST	2.97e+00
9	62	37.8	1616	1	P200_MYCGE	2.97e+00
10	61	37.2	390	1	RPSD_ANASP	4.44e+00
11	61	37.2	693	1	CAUP_DROME	4.44e+00
12	61	37.2	811	1	FS22_DROME	4.44e+00
13	61	37.2	873	1	FS21_DROME	4.44e+00
14	61	37.2	882	1	Y552_DROME	4.44e+00
15	60	36.6	294	1	NUSG_STRGR	6.62e+00
16	60	36.6	300	1	NUSG_STRCO	6.62e+00
17	60	36.6	404	1	ARRS_BOVIN	6.62e+00
18	60	36.6	455	1	INE2_CAEEL	6.62e+00
19	59	36.0	299	1	NUSG_STRVG	9.81e+00
20	59	36.0	309	1	NUSG_STRGB	9.81e+00
21	59	36.0	739	1	VNUC_EBOV	9.81e+00
22	59	36.0	877	1	PMT1_CANAL	9.81e+00
23	59	36.0	2137	1	SCPB_HUMAN	9.81e+00

24	59	36.0	3092	1	IRAL_YEAST	9.81e+00
25	58	35.4	330	1	LSPI_MOUSE	1.45e+01
26	58	35.4	348	1	RLA0_HALMA	1.45e+01
27	58	35.4	399	1	PHHC_PSEAE	1.45e+01
28	58	35.4	487	1	BERAMUNINE SYNTHASE (1.45e+01
29	58	35.4	674	1	DCMB_MOOTH	1.45e+01
30	58	35.4	773	1	FVB_MOUSE	1.45e+01
31	58	35.4	817	1	PPSA_PYRFU	1.45e+01
32	58	35.4	982	1	V120_HCMVA	1.45e+01
33	58	35.4	1234	1	MFD_MYCTU	1.45e+01
34	58	35.4	1894	1	FA52_YEAST	1.45e+01
35	57	34.8	326	1	ACCD_SYN3	2.12e+01
36	57	34.8	352	1	RLA0_HALCU	2.12e+01
37	57	34.8	352	1	RLA0_HALHA	2.12e+01
38	57	34.8	391	1	TAL_SYN3	2.12e+01
39	57	34.8	461	1	KLB2_ECOLI	2.12e+01
40	57	34.8	1319	1	DYNA_DROME	2.12e+01
41	56	34.1	219	1	RHIB_RHILV	3.09e+01
42	56	34.1	334	1	TAL1_YEAST	3.09e+01
43	56	34.1	341	1	CHFB_RAT	3.09e+01
44	56	34.1	431	1	ACEA_CORGL	3.09e+01
45	56	34.1	1113	1	N116_YEAST	3.09e+01

ALIGNMENTS

RESULT 1
 ID GCP_CAEEL STANDARD; PRT; 1005 AA.
 AC Q10663;
 DT 15-FEB-2000 (Rel. 39, Created)
 DT 15-FEB-2000 (Rel. 39, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE BIFUNCTIONAL GLYOXYLATE CYCLE PROTEIN [INCLUDES: ISOCITRATE LYASE
 DE (EC 4.1.3.1) (ISOCITRASE) (ISOCITRATASE) (ICL); MALATE SYNTHASE
 DE (EC 4.1.3.2)]
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 469-501.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE; 95301085.
 RA Liu F., Thatcher J.D., Barral J.M., Epstein H.F.;
 RT "Bifunctional glyoxylate cycle protein of Caenorhabditis elegans: a
 RT developmentally regulated, protein of intestine and muscle.";
 RL Dev. Biol. 169:399-414 (1995).
 CC -!- CATALYTIC ACTIVITY: ISOCITRATE = SUCCINATE + GLYOXYLATE.
 CC -!- CATALYTIC ACTIVITY: L-MALATE + COA = ACETYL-COA + H(2)O +
 CC GLYOXYLATE.
 CC -!- PATHWAY: FIRST AND SECOND STEP IN GLYOXYLATE BYPASS. AN
 CC ALTERNATIVE TO THE TRICARBOXYLIC ACID CYCLE (IN BACTERIA, PLANTS,
 CC AND FUNGI).
 CC -!- TISSUE SPECIFICITY: INTESTINAL AND BODY WALL MUSCLE CELLS.
 CC -!- DEVELOPMENTAL STAGE: FIRST DETECTED EARLY IN THE GENERATION OF
 CC THE CLONAL E CELL LINEAGE THAT IS COMMITTED TO INTESTINAL
 CC DIFFERENTIATING INTESTINAL AND BODY WALL MUSCLE CELLS. IN ACTIVELY
 CC HIGHEST ACTIVITY IN EMBRYOS AND FALLS DRAMATICALLY DURING L1
 CC LARVAL DEVELOPMENT.
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE ISOCITRATE
 CC LYASE FAMILY.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE MALATE
 CC SYNTHASE FAMILY.
 CC -!- CAUTION: IT IS UNCERTAIN IF MET-1 OR MET-40 IS THE INITIATOR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch)

```

DR EMBL: U23159; AAA85857.1; -.
DR PROSITE; PS00161; ISOCITRATE_LYASE; 1.
DR PROSITE; PS00510; MALATE_SYNTHASE; FALSE_NEG.
DR PFAM; PF00463; ICL; 2.
DR PFAM; PF01274; Malate_synthase; 1.
KW Glyoxylate bypass; Tricarboxylic acid cycle; Lyase;
KW Multifunctional enzyme.
FT DOMAIN 41 483 ISOCITRATE_LYASE.
FT DOMAIN 484 1005 MALATE_SYNTHASE.
FT ACT_SITE 230 230 PROBABLE.
SQ SEQUENCE 1005 AA; 112868 MW; 8C6D7D79310FD981 CRC64;

Query Match 40.9%; Score 67; DB 1; Length 1005;
Best Local Similarity 36.8%; Pred. No. 3.68e-01;
Matches 7; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

Db 266 ARTDAESSRLTSDIDPRD 284
QY 317 AKTEASPRMLDPQTEKE 335

RESULT 2
ID ACRA_MYCLE STANDARD; PRT; 606 AA.
AC P46831;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ISOCITRATE LYASE (EC 4.1.3.1) (ISOCITRASE) (ISOCITRATASE) (ICL).
GN ACEA.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96059637.
RA Fshihi H., Cole S.T.;
RT "The Mycobacterium leprae genome: systematic sequence analysis
RT identifies key catabolic enzymes, ATP-dependent transport systems and
RT a novel pola locus associated with genomic variability.";
RL Mol. Microbiol. 16:909-919(1995).
CC -!- CATALYTIC ACTIVITY: ISOCITRATE = SUCCINATE + GLYOXYLATE.
CC -!- COFACTOR: REQUIRES DIVALENT CATIONS (BY SIMILARITY).
CC -!- PATHWAY: FIRST STEP IN GLYOXYLATE BYPASS, AN ALTERNATIVE TO THE
CC TRICARBOXYLIC ACID CYCLE (IN BACTERIA, PLANTS, AND FUNGI).
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE ISOCITRATE LYASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 246257; CAA86357.1; -.
DR PROSITE; PS00161; ISOCITRATE_LYASE; 1.
DR PFAM; PF00463; ICL; 2.
KW Glyoxylate bypass; Tricarboxylic acid cycle; Lyase; Phosphorylation.
FT ACT_SITE 211 211 BY SIMILARITY.
SQ SEQUENCE 606 AA; 67571 MW; 0111A835D8F5C7F2 CRC64;

Query Match 40.28; Score 66; DB 1; Length 606;
Best Local Similarity 31.68; Pred. No. 5.63e-01;
Matches 6; Conservative 12; Mismatches 1; Indels 0; Gaps 0;

Db 247 ARTDAPAAANLSDRDERD 265
QY 317 AKTEASPRMLDPQTEKE 335

RESULT 3

```

```

ID AROB_BACSU STANDARD; PRT; 362 AA.
AC P31102;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 3-DEHYDROQUINATE SYNTHASE (EC 4.6.1.3).
GN AROB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RA Henner D.J.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: 7-PHOSPHO-3-DEOXY-ARABINO-HEPTULOSONATE =
CC 3-DEHYDROQUINATE + ORTHOPHOSPHATE.
CC -!- PATHWAY: SECOND STEP IN THE BIOSYNTHESIS FROM CHORISMATE OF THE
CC AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M80245; AAA20860.1; -.
DR EMBL; Z99115; CAB4186.1; -.
DR SUBTILIST; BGI0285; AROB.
KW Aromatic amino acid biosynthesis; Lyase.
SQ SEQUENCE 362 AA; 40817 MW; A0093C7730916AAC CRC64;

Query Match 38.4%; Score 63; DB 1; Length 362;
Best Local Similarity 28.0%; Pred. No. 1.97e+00;
Matches 7; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

Db 305 RKETSTVLLNRMMNDKKTGKQI 329
QY 318 KTEASPRMLDPQTEKEASGSLQ 342

RESULT 4
ID HMBL_SOYBN STANDARD; PRT; 379 AA.
AC P46608;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HOMEBOX PROTEIN SBH1.
GN H1.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eufrosids I; Fabales; Fabaceae; Papilionoideae;
OC Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SOMATIC EMBRYO;
RX MEDLINE; 94169300.
RA Ma H., McMullen M.D., Finer J.J.;
RT "Identification of a homeobox-containing gene with enhanced
RT expression during soybean (Glycine max L.) somatic embryo
RT development.";
RL Plant Mol. Biol. 24:465-473(1994).
CC -!- FUNCTION: POSSIBLE TRANSCRIPTION ACTIVATOR INVOLVED IN EARLY
CC EMBRYONIC DEVELOPMENT. PROBABLY BINDS TO THE DNA SEQUENCE 5'-TGAC-
CC 3'.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- TISSUE SPECIFICITY: EXPRESSED MAINLY IN EMBRYONIC TISSUES. WEAKLY
CC DETECTED IN STEMS AND HYPOCOTYL.
CC -!- DEVELOPMENTAL STAGE: FIRST EXPRESSED IN THE EMBRYO PROLIFERATION

```


STAGE, INCREASES DURING EARLY SOMATIC EMBRYO DEVELOPMENT AND DECREASES THEREAFTER.
-!- SIMILARITY: BELONGS TO THE TALE/KNOX FAMILY OF HOMEBOX PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; LI3663; AAA20882.1; -
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR PFAM; PF00046; homeobox; 1.
KW Homeobox; DNA-binding; Nuclear protein; Transcription regulation;
KW Activator.
FT DOMAIN 102 107 POLY-HIS.
FT DOMAIN 109 113 POLY-ASN.
FT DOMAIN 115 122 POLY-SER.
FT DOMAIN 244 248 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 259 262 ELK DOMAIN.
FT DNA_BIND 283 345 HOMEBOX (TALE-TYPE).
SQ SEQUENCE 379 AA; 42374 MW; 8865020029EDD5E4 CRC64;
Query Match 38.4%; Score 63; DB 1; Length 379;
Best Local Similarity 30.4%; Pred. No. 1.97e+00;
Matches 7; Conservative 8; Mismatches 8; Indels 0; Gaps 0;
Db 246 DVDLHMIDPQAEEDRLKQLLR 268
QY 320 EAESPRMLDPQTEKEASGSLQ 342

RESULT 5
ID MRXK_KLEPN STANDARD; PRT; 828 AA.
AC P21647;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE OUTER MEMBRANE USHER PROTEIN MRKC PRECURSOR.
GN MRKC.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
FN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-IA565;
RX MEDLINE; 91100388.
RA Allen B.L., Gerlach G.-F., Clegg S.;
RT "Nucleotide sequence and functions of mrk determinants necessary for expression of type 3 fimbriae in Klebsiella pneumoniae."
RL J. Bacteriol. 173:916-920(1991).
CC -!- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF THE TYPE 3 FIMBRIAL SUBUNIT (MRKA).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M55912; AAA25095.1; -
DR PIR; D39142; D39142.
DR PROSITE; PS01151; FIMBRIAL_USHER; 1.
DR PFAM; PF00577; Usher; 1.
KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.

FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 828 OUTER MEMBRANE USHER PROTEIN MRKC.
FT DISULFID 813 827 POTENTIAL.
SQ SEQUENCE 828 AA; 91049 MW; B30EDF5798249FC9 CRC64;
Query Match 38.4%; Score 63; DB 1; Length 828;
Best Local Similarity 55.0%; Pred. No. 1.97e+00;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
Db 106 SPEKLNPSLSEKSGFCGRILQ 125
QY 323 SPRMLDPQTEKEASGSLQ4342

RESULT 6
ID TCOF_HUMAN STANDARD; PRT; 1411 AA.
AC Q13428; Q99408;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE TREACLE PROTEIN (TREACHER COLLINS SYNDROME PROTEIN).
GN TCOF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96154183.
RA Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Loftus S.K., Bonner C.A., Koprivnikar K., Wasmuth J.J.;
RT "Positional cloning of a gene involved in the pathogenesis of Treacher Collins syndrome. The Treacher Collins Syndrome Collaborative Group."
RL Nat. Genet. 12:130-136(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97250498.
RA Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M., Ashley J.A., Lovett M., Jabs E.W.;
RT "TCOF1 gene encodes a putative nucleolar phosphoprotein that exhibits mutations in Treacher Collins Syndrome throughout its coding region."
RL Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).
RN [3]
RP VARIANTS L-439; V-810; V-1313 & G-1355, AND VARIANT TCS R-53.
RX MEDLINE; 97195537.
RA Edwards S.J., Gladwin A.J., Dixon M.J.;
RT "The mutational spectrum in Treacher Collins syndrome reveals a predominance of mutations that create a premature-termination codon."
RL Am. J. Hum. Genet. 60:515-524(1997).
CC -!- DISEASE: DEFECTS IN TCOF1 ARE THE CAUSE OF TREACHER COLLINS SYNDROME (TCS). TCS IS A AUTOSOMAL DOMINANT DISORDER OF CRANIOFACIAL DEVELOPMENT THAT OCCURS WITH AN INCIDENCE OF 1/50,000 LIVE BIRTHS. THE CLINICAL FEATURES OF TCS ARE BILATERALLY SYMMETRICAL AND INCLUDE: (1) ABNORMALITIES OF THE EXTERNAL EARS, ATRESIA OF THE EXTERNAL EAR CANALS, AND MALFORMATION OF THE MIDDLE EAR OSSICLES, WHICH MAY RESULT IN CONDUCTIVE HEARING LOSS; (2) LATERAL DOWNWARD SLOPING OF PALPEBRAL FISSURES, FREQUENTLY WITH COLOBOMAS OF THE LOWER EYELIDS; (3) HYPOPLASIA OF THE MANDIBLE AND ZYGOMATIC COMPLEX; (4) CLEFT PALATE.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U40847; AAC50903.1; -
DR EMBL; U76366; AAC51181.1; -
DR EMBL; U84664; AAC51185.1; -

DR EMBL: U84640; AAC51185.1; JOINED.
DR EMBL: U84641; AAC51185.1; JOINED.
DR EMBL: U84642; AAC51185.1; JOINED.
DR EMBL: U84643; AAC51185.1; JOINED.
DR EMBL: U84644; AAC51185.1; JOINED.
DR EMBL: U84645; AAC51185.1; JOINED.
DR EMBL: U84646; AAC51185.1; JOINED.
DR EMBL: U84647; AAC51185.1; JOINED.
DR EMBL: U84648; AAC51185.1; JOINED.
DR EMBL: U84649; AAC51185.1; JOINED.
DR EMBL: U84650; AAC51185.1; JOINED.
DR EMBL: U84651; AAC51185.1; JOINED.
DR EMBL: U84652; AAC51185.1; JOINED.
DR EMBL: U84653; AAC51185.1; JOINED.
DR EMBL: U84654; AAC51185.1; JOINED.
DR EMBL: U84655; AAC51185.1; JOINED.
DR EMBL: U84656; AAC51185.1; JOINED.
DR EMBL: U84657; AAC51185.1; JOINED.
DR EMBL: U84658; AAC51185.1; JOINED.
DR EMBL: U84659; AAC51185.1; JOINED.
DR EMBL: U84660; AAC51185.1; JOINED.
DR EMBL: U84661; AAC51185.1; JOINED.
DR EMBL: U84662; AAC51185.1; JOINED.
DR EMBL: U84663; AAC51185.1; JOINED.
DR MIM: 154500; -.
KW Disease mutation; Polymorphism.
FT DOMAIN 89 97 POLY-GLU.
FT DOMAIN 204 207 POLY-SER.
FT DOMAIN 616 619 POLY-SER.
FT DOMAIN 919 924 POLY-SER.
FT DOMAIN 1285 1289 POLY-LYS.
FT DOMAIN 1375 1386 POLY-LYS.
FT DOMAIN 1398 1405 POLY-LYS.
FT VARIANT 53 53 W -> R (IN TCS).
FT VARIANT 439 439 /FTid=VAR_005630.
FT VARIANT 810 810 /FTid=VAR_005631.
FT VARIANT 1313 1313 /FTid=VAR_005632.
FT VARIANT 1355 1355 /FTid=VAR_005633.
FT VARIANT 1355 1355 /FTid=VAR_005634.
FT SEQUENCE 1411 AA; 144312 MW; 3880203D985C2699 CRC64;
Query Match 38.4%; Score 63; DB 1; Length 1411;
Best Local Similarity 41.7%; Pred. No. 1.97e+00;
Matches 10; Conservative 7; Mismatches 6; Indels 1; Gaps 1;
Db 1170 AKPO-QAAGMLSPKTKGKAASGT 1192
QY 317 AKTEASPRMLDPQTEKEASGGS 340
RESULT 7
ID PAP2_XENLA STANDARD; PRT; 484 AA.
AC P51005;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE POLY(A) POLYMERASE TYPE 2 (EC 2.7.7.19) (PAP) (POLYNUCLEOTIDE
ADENYLITRANSFERASE).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae;
OC Xenopus.
RN [1].
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RX MEDLINE; 96079940.
RA Ballantyne S., Bilger A., Astrom J., Virtanen A., Wickens M.;
RT "Poly (A) polymerases in the nucleus and cytoplasm of frog oocytes;
dynamic changes during oocyte maturation and early development.";

RL RNA 1:64-78(1995).
-1- FUNCTION: POLYMERASE THAT CREATES THE 3' POLY(A) TAIL OF MRNA'S.
CC MAY ACQUIRE SPECIFICITY THROUGH INTERACTION WITH A CLEAVAGE AND
CC POLYADENYLATION FACTOR (CPSF).
CC -1- CATALYTIC ACTIVITY: N ATP + (NUCLEOTIDE)(M) - N PYROPHOSPHATE +
CC (NUCLEOTIDE)(M+N).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U19974; AAC59746.1; -.
DR MRNA processing; Transferase; Transcription; RNA-binding;
KW Nuclear protein.
FT DOMAIN 240 257 NUCLEAR LOCALIZATION SIGNAL 1 (BY
FT SIMILARITY).
FT DOMAIN 392 407 NUCLEAR LOCALIZATION SIGNAL 2 (BY
FT SIMILARITY).
FT SEQUENCE 484 AA; 53830 MW; D81361CFOD79B879 CRC64;
Query Match 37.8%; Score 62; DB 1; Length 484;
Best Local Similarity 39.1%; Pred. No. 2.97e+00;
Matches 9; Conservative 7; Mismatches 7; Indels 0; Gaps 0;
Db 317 TNOATDVMVPNNSTENSGSL 339
QY 319 TEAESPRMLDPQTEKEASGSL 341
RESULT 8
ID MSH1_YEAST STANDARD; PRT; 959 AA.
AC P25846;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MUTS PROTEIN HOMOLOG 1.
GN MSH1 OR YHR120W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1].
RP SEQUENCE FROM N.A.
RX MEDLINE; 93093456.
RA Reenan R.A.G., Kolodner R.D.;
RT "Isolation and characterization of two Saccharomyces cerevisiae genes
encoding homologs of the bacterial HexA and Muts mismatch repair
proteins.";
RL Genetics 132:963-973(1992).
RN [2].
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RX MEDLINE; 94378003.
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
Du Z., Favella A., Fulton L., Gattung S., Geisel C., Kirsten J.,
Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
VIII.";
RN [3].
RP CHARACTERIZATION.
RX MEDLINE; 93093457.
RA Reenan R.A.G., Kolodner R.D.;
RT "Characterization of insertion mutations in the Saccharomyces

RT cerevisiae MSH1 and MSH2 genes: evidence for separate mitochondrial
 RL and nuclear functions.";
 CC -!- FUNCTION: INVOLVED IN MITOCHONDRIAL DNA REPAIR.
 CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 CC EMBL; M84169; AAA34801.1; -;
 DR EMBL; U00059; AAB68849.1; -;
 DR PIR; S27432; S27432;
 DR PIR; S48962; S48962;
 DR SGD; L000189; MSH1.
 DR PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
 DR PFAM; PF00488; muts; 1.
 DR PFAM; PF01624; muts; 1.
 KW DNA repair; ATP-binding; DNA-binding; Mitochondrion.
 FT NP_BIND 771 778 ATP (POTENTIAL).
 FT CONFLICT 437 437 T -> N (IN REF. 1).
 FT CONFLICT 695 695 E -> V (IN REF. 1).
 SQ SEQUENCE 959 AA; 109408 MW; F40985CB4D2ECFF1 CRC64;
 Query Match 37.8%; Score 62; DB 1; Length 959;
 Best Local Similarity 42.9%; Pred. No. 2.97e+00;
 Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
 Db 529 AKQNAVTRMLDIDVKKES 549
 QY 317 AKTEASPRMLDPQTEKEAS 337
 RESULT 9
 ID P200_MYCGE STANDARD; PRT; 1616 AA.
 AC Q49429; Q49259; Q49298; Q49352; Q49353;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PROTEIN P200.
 GN MG386.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE; 96026346.
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
 RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.F., Dougherty B.A., Bitt K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RT "The minimal gene complement of Mycoplasma genitalium.";
 RL Science 270:397-403(1995).
 RN [2]
 RP SEQUENCE OF 256-427; 432-543 AND 1083-1140 FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE; 94075230.
 RA Peterson S.N., Hu P.-C., Bitt K.F., Hutchison C.A. III;
 RT "A survey of the Mycoplasma genitalium genome by using random
 RT sequencing.";
 RL J. Bacteriol. 175:7918-7930(1993).
 CC -!- FUNCTION: COULD BE A ACCESSORY STRUCTURAL COMPONENT IN
 CC CYTADHERENCE (BY SIMILARITY).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 CC EMBL; U39720; AAC71613.1; -;
 DR EMBL; U02245; AAA03400.1; -;
 DR EMBL; U02245; AAA03401.1; ALT_FRAME.
 DR EMBL; U02175; AAD12458.1; -;
 DR EMBL; U02126; AAD12402.1; -;
 DR TIGR; MG386; -;
 KW Cytadherence; Structural protein; Repeat.
 FT DOMAIN 1205 1389 2 X 32 AA REPEAT.
 FT REPEAT 1205 1236 2-1.
 FT REPEAT 1358 1389 2-2.
 FT DOMAIN 891 1389 2 X 26 AA REPEAT.
 FT REPEAT 1161 1186 2-1.
 FT REPEAT 1310 1339 2-2.
 FT CONFLICT 256 256 P -> S (IN REF. 2).
 FT CONFLICT 304 304 S -> F (IN REF. 2).
 SQ SEQUENCE 1616 AA; 185678 MW; 6AF76A13AC49E4FF CRC64;
 Query Match 37.8%; Score 62; DB 1; Length 1616;
 Best Local Similarity 42.9%; Pred. No. 2.97e+00;
 Matches 9; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
 Db 803 ESNALFSDKDEKSSDSELQ 823
 QY 322 ESRMLDPQTEKEASGSLQ 342
 RESULT 10
 ID RPSD_ANASP STANDARD; PRT; 390 AA.
 AC P26683;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE RNA POLYMERASE SIGMA FACTOR RPOD (SIGMA A).
 GN RPOD OR SIGA.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE; 91193199.
 RA Brahamsha B., Haselkorn R.;
 RT "Isolation and characterization of the gene encoding the principal
 RT sigma factor of the vegetative cell RNA polymerase from the
 RT cyanobacterium Anabaena sp. strain PCC 7120.";
 RL J. Bacteriol. 173:2442-2450(1991).
 CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
 CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
 CC THEN IS RELEASED. THIS IS THE PRIMARY SIGMA-FACTOR OF THIS
 CC BACTERIA.
 CC
 CC -!- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 CC EMBL; M60046; AAA22043.1; -;
 DR PIR; A42724; A42724.
 DR HSP; P00579; ISIG.
 DR PRINTS; PR00046; SIGMA70FCT.
 DR PROSITE; PS00715; SIGMA70_1; 1.
 DR PROSITE; PS00716; SIGMA70_2; 1.
 DR PFAM; PF00140; sigma70; 1.
 KW Transcription regulation; Sigma factor; DNA-directed RNA polymerase;
 KW DNA-binding.

```

FT DOMAIN 182 195 POLYMERASE CORE BINDING (POTENTIAL).
FT DNA_BIND 351 370 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 390 AA: 45641 MW; 604814660815284D CRC64;

Query Match 37.2%; Score 61; DB 1; Length 390;
Best Local Similarity 35.0%; Pred. No. 4.44e+00;
Matches 7; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Db 42 DGDIDFELEPQTDDEDAKSG 61
   ::: |::|::|::|::|
QY 320 EAESPRMLDPQTEKEASGG 339

RESULT 11
ID CAUP_DROME STANDARD; PRT; 693 AA.
AC P54269;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HOMEOBOX PROTEIN CAUPOLICAN.
GN CAUP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96180722.
RA Gomez-Skarmeta J.-L., del Corral R.D., de la Calle-Mustienes E.,
RA Ferras-Marco D., Modolell J.;
RT "Araucan and caupolican, two members of the novel iroquois complex,
RT encode homeoproteins that control proneural and vein-forming genes.";
RL Cell 85:95-110(1996).
CC -!- FUNCTION: CONTROLS PRONEURAL AND VEIN FORMING GENES. POSITIVE
CC TRANSCRIPTIONAL CONTROLLER OF AC-SC (ACHAETE-SCUTE). MAY ACT AS AN
CC ACTIVATOR THAT INTERACTS WITH THE TRANSCRIPTIONAL COMPLEX
CC ASSEMBLED ON THE AC AND SC PROMOTERS AND PARTICIPATES IN
CC TRANSCRIPTION INITIATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE TALE/IRO FAMILY OF HOMEOBOX PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X95178; CAA64485.1; -.
CC HSSP; P02833; 1SAN.
CC DR FLYBASE; FBgn0015919; caup.
CC DR PROSITE; PS00027; HOMEOBOX_1; 1.
CC DR PROSITE; PS50071; HOMEOBOX_2; 1.
CC DR PFAM; PF00046; homeobox; 1.
CC KW Transcription regulation; DNA-binding; Homeobox; Nuclear protein;
CC Developmental protein.
CC FT DNA_BIND 226 288 HOMEOBOX (TALE-TYPE).
CC FT DOMAIN 300 303 POLY-ASP.
CC FT DOMAIN 405 418 POLY-GLN.
CC FT DOMAIN 501 516 POLY-GLN.
CC FT DOMAIN 517 528 POLY-HIS.
CC FT DOMAIN 565 572 POLY-SER.
CC FT DOMAIN 613 624 POLY-SER.
CC SEQUENCE 693 AA; 73749 MW; 8E0D6D43C9CDC619 CRC64;

Query Match 37.2%; Score 61; DB 1; Length 693;
Best Local Similarity 40.0%; Pred. No. 4.44e+00;
Matches 10; Conservative 5; Mismatches 9; Indels 1; Gaps 1;

Db 297 KTEDDDGMMSDDEKEKDAADGK 321
   |||::|::|::|::|
QY 318 KTEAESPRMLDPQTEKEAS-GGSL 341

```

```

RESULT 12
ID FS22_DROME STANDARD; PRT; 811 AA.
AC P34083;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE FASCICLIN II, PHOSPHATIDYLINOSITOL-LINKED ISOFORM PRECURSOR (FAS II).
GN FAS2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON-S;
RX MEDLINE; 92005695.
RA Greeningloh G., Rehm E.J., Goodman C.S.;
RT "Genetic analysis of growth cone guidance in Drosophila: fasciclin II
RT functions as a neuronal recognition molecule.";
RL Cell 67:45-57(1991).
CC -!- FUNCTION: NEURONAL RECOGNITION MOLECULE. INVOLVED IN A PATHWAY
CC RECOGNITION FOR AXONS DURING THE DEVELOPMENT OF NERVE FASCICLES.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- ALTERNATIVE PRODUCTS: THE PHOSPHATIDYLINOSITOL-LINKED AND THE
CC TRANSMEMBRANE ISOFORMS ARE GENERATED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: INITIALLY EXPRESSED ON THE SURFACE OF THE
CC AXONS IN THE MP1 PATHWAY (AND LATER ON SEVERAL OTHER LONGITUDINAL
CC AXON FASCICLES).
CC -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M77165; AAA28527.1; -.
CC PIR; A41054; A41504.
CC DR PIR; A41054; A41054.
CC DR FLYBASE; FBgn000635; Fas2.
CC DR PFAM; PF00041; fn3; 2.
CC DR PFAM; PF00047; ig; 4.
CC KW Cell adhesion; Glycoprotein; Repeat; Alternative splicing;
CC Immunoglobulin domain; GPI-anchor; Signal; Neurogenesis.
CC
CC SIGNAL 1 28
CC FT CHAIN 29 811 POTENTIAL.
CC FT DOMAIN 47 123 FASCICLIN II, GPI-LINKED ISOFORM.
CC FT DOMAIN 152 214 IG-LIKE C2-TYPE DOMAIN.
CC FT DOMAIN 244 309 IG-LIKE C2-TYPE DOMAIN.
CC FT DOMAIN 336 414 IG-LIKE C2-TYPE DOMAIN.
CC FT DOMAIN 444 511 IG-LIKE C2-TYPE DOMAIN.
CC FT DOMAIN 544 619 FIBRONECTIN TYPE-III.
CC FT DOMAIN 648 705 FIBRONECTIN TYPE-III.
CC FT DISULFID 54 116 POTENTIAL.
CC FT DISULFID 159 207 POTENTIAL.
CC FT DISULFID 251 302 POTENTIAL.
CC FT DISULFID 343 407 POTENTIAL.
CC FT DISULFID 451 504 POTENTIAL.
CC FT CARBOHYD 74 74 POTENTIAL.
CC FT CARBOHYD 250 250 POTENTIAL.
CC FT CARBOHYD 330 330 POTENTIAL.
CC FT CARBOHYD 448 448 POTENTIAL.
CC FT CARBOHYD 458 458 POTENTIAL.
CC FT CARBOHYD 576 576 POTENTIAL.
CC SEQUENCE 811 AA; 90067 MW; FAB06B4D2ECC57B9 CRC64;

Query Match 37.2%; Score 61; DB 1; Length 811;
Best Local Similarity 36.0%; Pred. No. 4.44e+00;

```

STRAIN=N2-3-11;
MEDLINE: 94314177

STRAIN=N2-3-11;
MEDLINE: 94314177

Thu May 11 06:50:18 2000

```

RA Kuberski S., Kasberg T., Distler J.;
RT "The nusG gene of Streptomyces griseus: cloning of the gene and
RT analysis of the A-factor binding properties of the gene product.";
RL FEMS Microbiol. Lett. 119:33-39(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-IFO 13350;
RX MEDLINE; 94114580.
RA Miyake K., Onaka H., Horinouchi S., Beppu T.;
RT "Organization and nucleotide sequence of the secE-nusG region of
RT Streptomyces griseus.";
RL Biochim. Biophys. Acta 1217:97-100(1994).
CC -!- FUNCTION: INFLUENCES TRANSCRIPTION TERMINATION AND
CC ANTITERMINATION. ACTS AS A COMPONENT OF THE TRANSCRIPTION COMPLEX,
CC AND INTERACTS WITH THE TERMINATION FACTOR RHO AND RNA POLYMERASE
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUSG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X72787; CAA51296.1; -.
DR EMBL; D17464; BAA04281.1; -.
DR PIR; S32234; S32234.
DR HSP; 002248; 2BCT.
DR PRINTS; PR00338; NUSGTNSCPFCT.
DR PROSITE; PS01014; NUSG; 1.
KW Transcription termination.
FT CONFLICT 84 88 AEPAA -> RAGRR (IN REF. 2).
FT CONFLICT 285 286 EL -> DV (IN REF. 2).
SQ SEQUENCE 294 AA; 31810 MW; C51CD316DB4977F5 CRC64;

Query Match 36.6%; Score 60; DB 1; Length 294;
Best Local Similarity 43.8%; Pred. No. 6.62e+00;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

DB 207 KMLAPEAEKAEKAREAA 222
QY 325 RMLDPQTEKEKESGGS 340

Search completed: Wed May 10 14:08:54 2000
Job time : 102 secs.

```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Wed May 10 14:09:12 2000; MasPar time 227.63 Seconds

Tabular output not generated. 7.919 Million cell updates/sec

Title: >US-09-376-430-2
Description: (317-342) from US09376430A.pep (23 of 25)
Perfect Score: 164
Sequence: 1 AKTEAESPRMLDPQTEKEASGGSLQ 26

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database: sptrembl12

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 28.180; Variance 44.093; scale 0.639

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	67	40.9	526	C05E4.9 PROTEIN (FRAGM	1.44e+00
2	66	40.2	367	ACEA.	2.13e+00
3	66	40.2	424	FORK HEAD DOMAIN PROTE	2.13e+00
4	66	40.2	1095	PROTEIN TYROSINE KINAS	2.13e+00
5	65	39.6	485	VIRE3 PROTEIN (FRAGMEN	3.14e+00
6	63	38.4	214	XPOL-141.	6.76e+00
7	63	38.4	506	HYPOTHETICAL 55.0 KD P	6.76e+00
8	63	38.4	1411	TREACHER COLLINS SYND	9.86e+00
9	62	37.8	2346	TPR HOMOLOG.	1.43e+01
10	61	37.2	249	TWEAK.	1.43e+01
11	61	37.2	371	KNOTTED I CLASS HOMEOD	1.43e+01
12	61	37.2	375	LUNATIC FRINGE.	1.43e+01
13	61	37.2	829	GLYCOPROTEIN B.	1.43e+01
14	61	37.2	829	GLYCOPROTEIN B.	1.43e+01
15	60	36.6	236	HYPOTHETICAL 24.0 KD P	2.07e+01
16	60	36.6	275	F40F9.3 PROTEIN.	2.07e+01
17	60	36.6	321	CHROMOSOME XII COSMID	2.07e+01
18	60	36.6	1239	CHROMOSOME XII COSMID	2.07e+01
19	60	36.6	1302	NEUROGLIAN.	2.07e+01
20	60	36.6	1616	KIAA0346 PROTEIN (FRAG	2.07e+01

21	60	36.6	1684	4	000310	GRAVIN.	2.07e+01
22	60	36.6	1741	4	000498	MYASTHENIA GRAVIS AUTO	2.07e+01
23	60	36.6	1781	4	099970	GRAVIN.	2.07e+01
24	60	36.6	2489	3	061116	CHROMOSOME XVI COSMID	2.07e+01
25	59	36.0	178	2	060074	IMMATURE FIMBRILLIN PR	2.99e+01
26	59	36.0	399	4	014726	BETA-SPECTRIN (FRAGMEN	2.99e+01
27	59	36.0	739	14	09YMG4	NUCLEOPROTEIN.	2.99e+01
28	59	36.0	739	14	072142	NUCLEOCAPSID PROTEIN.	2.99e+01
29	59	36.0	1280	13	Q90933	NEURON-GLIA CELL ADHES	2.99e+01
30	58	35.4	122	11	Q62024	PP52 (FRAGMENT).	4.28e+01
31	58	35.4	124	11	Q62023	PP52 PROTEIN (FRAGMENT	4.28e+01
32	58	35.4	330	11	Q62022	P50B.	4.28e+01
33	58	35.4	424	14	Q65148	PEP424R.	4.28e+01
34	58	35.4	428	2	P95328	ISOCITRATE LYASE.	4.28e+01
35	58	35.4	616	10	Q9XIJ3	T10024.21.	4.28e+01
36	58	35.4	641	5	Q61884	F59B1.9 PROTEIN.	4.28e+01
37	58	35.4	714	2	Q60149	DNA TOPOISOMERASE I (E	4.28e+01
38	58	35.4	819	11	Q922H3	RYN BINDING PROTEIN FY	4.28e+01
39	58	35.4	1027	5	Q09615	HYPOTHETICAL 115.8 KD	4.28e+01
40	58	35.4	1106	5	Q97207	L4830.1 PROTEIN.	4.28e+01
41	58	35.4	1111	2	Q86522	HYPOTHETICAL 119.9 KD	4.28e+01
42	58	35.4	1181	10	Q04650	SIMILARITY TO TROPOMYO	4.28e+01
43	58	35.4	1584	4	Q14514	BAI 1.	4.28e+01
44	58	35.4	1887	3	Q12533	FAS2 PROTEIN.	4.28e+01
45	58	35.4	3896	14	P87514	PESTIVIRUS POLYPROTEIN	4.28e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	526 AA.
ID	017353			
AC	017353;			
DT	01-JAN-1998 (TrEMBLrel. 05, Created)			
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)			
DT	01-NOV-1998 (TrEMBLrel. 08, Last annotation update)			
DE	C05E4.9 PROTEIN (FRAGMENT).			
GN	C05E4.9			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;			
OC	Rhabditina; Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN-BRISTOL N2;			
RX	MEDLINE; 94150718;			
RA	WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,			
RA	BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., FULTON L.,			
RA	CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,			
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,			
RA	JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,			
RA	LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,			
RA	PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,			
RA	SMALDON N., SMITH A., SONNHAMMER E., STADEN K., SULSTON J.,			
RA	THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,			
RA	WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,			
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.			
RT	elegans";			
RL	Nature 368:32-38(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RA	BLANCHARD M., KRAMER J., GIBSON A.,			
RL	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RA	WATERSTON R.,			
RL	Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF024209; AAB71278.1;			
DR	PFAM; PF00463; ICL; 2.			
FT	NON_PER 526			
SQ	SEQUENCE 526 AA; 58642 MW; 736CC90A CRC32;			

Query Match 40.9%; Score 67; DB 5; Length 526;

DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PFAM; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
SQ SEQUENCE 424 AA; 47122 MW; C5006797 CRC32;

Query Match 40.2%; Score 66; DB 13; Length 424;
Best Local Similarity 34.6%; Pred. No. 2.13e+00;
Matches 9; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

Ddb 267 GKREQSSARALRPATDQOTRAHGLQ 292
:| | :| | | :| :| :| :|
QY 317 AKTEASPRMLDPQTKEEAGSGSLQ 342

RESULT 4 PRELIMINARY; PRT; 1095 AA.

ID	C93596
AC	Q93596;
DT	01-NOV-1998 (TrEMBLrel. 08, Created)
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE	01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE	PROTEIN TYROSINE KINASE (EC 2.7.1.112).
GN	JAK2A.
OS	Brachydanio rerio (Zebrafish) (Zebra danio).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC	Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC	Cyprinidae; Cyprininae; Rasborinae; Danio.
RN	[1]
RP	SEQUENCE FROM N.A.
RA	GATES A.C.;
RL	Thesis (1998), University of Melbourne, Australia.
DR	ENBL; AJ005690; CA006674.1; -.
DR	HSP; P08631; 2HCK.
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR	PFAM; PF00069; pkinase; 2.
KW	Transferase.
SQ	SEQUENCE 1095 AA; 125171 MW; 961AB4C1 CRC32;

Query Match 40.2%; Score 66; DB 13; Length 1095;
Best Local Similarity 50.0%; Pred. No. 2.13e+00;
Matches 9; Conservative 6; Mismatches 2; Indels 1; Gaps 1

Ddb 2 MLANTENEOKQASCAALQ 19
|| : : || :| :| :| :|
QY 326 ML-DPQTEEKESGGSGLQ 342

RESULT 5 PRELIMINARY; PRT; 485 AA.

ID	Q9WB7
AC	Q9WB7;
DT	01-NOV-1999 (TrEMBLrel. 12, Created)
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE	01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE	VIRE3 PROTEIN (FRAGMENT).
OS	Agrobacterium tumefaciens.
OG	Plasmid Ti.
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC	Rhizobiaceae; Agrobacterium.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C58;
RA	KALOGERAKI V.S.;
ET	"Fragment of the vire3 gene of Agrobacterium tumefaciens.";
RL	Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AF153204; AAD34168.1; -.
DR	AF153204; AAD34168.1; -.
KW	Plasmid.
FT	NON_TER
SQ	SEQUENCE 485 AA; 53835 MW; BA97524D CRC32;

Query Match 39.6%; Score 65; DB 2; Length 485;
Best Local Similarity 40.7%; Pred. No. 3.14e+00;
Matches 11; Conservative 9; Mismatches 5; Indels 2; Gaps 2


```

Db 459 PKADHERVLESSRSRSMGSL 485
QY 317 AKTEASPRMLDP-QTEKEASGG-SL 341

RESULT 6
ID Q9X397 PRELIMINARY; PRT; 214 AA.
AC Q9X397;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PX01-141.
OS Bacillus anthracis.
OG Plasmid virulence plasmid PX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-STERNE;
RA OKINAKA R.T., CLOUD K., HAMTON O., HOFFMASTER A., HILL K.K., KEIM P.,
RA KOEHLER T., LAMKE G., KUMANO S., MAHILLON J., MANTER D., MARTINEZ Y.,
RA RICKS D.O., SVENSSON R., JACKSON P.J.;
RT "The sequence and organization of pX01, the large Bacillus anthracis
RT plasmid harboring the Anthrax toxin genes.";
RL J. Bacteriol. 0:0-0(1999).
DR EMBL; AF065404; RAD32444.1; -.
KW Plasmid.
SQ SEQUENCE 214 AA; 23913 MW; 2652D2F1 CRC32;

Query Match 38.4%; Score 63; DB 2; Length 214;
Best Local Similarity 50.0%; Pred. No. 6.76e+00;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 178 PEMLEKEKEAKAG 193
QY 324 PRMLDPQTEKEASGG 339

RESULT 7
ID P96824 PRELIMINARY; PRT; 506 AA.
AC P96824;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HYPOTHETICAL 55.0 KD PROTEIN.
GN MTC15.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA BADCOCK K., CHURCHER C.M.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA COLE S.T.;
RT "An integrated map of the genome of the tubercle bacillus,
RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT leprae.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR EMBL; Z92770; CAB07053.1; -.
DR HSSP; P11883; IAD3.
DR PFAM; PF00171; aldedh; 1.

KW Hypothetical protein.
SQ SEQUENCE 506 AA; 55035 MW; 4F7E4868 CRC32;

Query Match 38.4%; Score 63; DB 2; Length 506;
Best Local Similarity 42.1%; Pred. No. 6.76e+00;
Matches 8; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Db 28 KSETEAPREPAPVSDKQT 46
QY 318 KTEASPRMLDPQTEKEA 336

RESULT 8
ID Q99860 PRELIMINARY; PRT; 1411 AA.
AC Q99860;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
DE TREACHER COLLINS SYNDROME.
GN TCOF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA DIXON J., EDWARDS S.J., ANDERSON I., BRASS A., SCAMBLER P.J.,
RA DIXON M.J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U79659; AAB40722.1; -.
DR EMBL; U79645; AAB40722.1; JOINED.
DR EMBL; U79646; AAB40722.1; JOINED.
DR EMBL; U79647; AAB40722.1; JOINED.
DR EMBL; U79648; AAB40722.1; JOINED.
DR EMBL; U79649; AAB40722.1; JOINED.
DR EMBL; U79650; AAB40722.1; JOINED.
DR EMBL; U79651; AAB40722.1; JOINED.
DR EMBL; U79652; AAB40722.1; JOINED.
DR EMBL; U79653; AAB40722.1; JOINED.
DR EMBL; U79654; AAB40722.1; JOINED.
DR EMBL; U79655; AAB40722.1; JOINED.
DR EMBL; U79656; AAB40722.1; JOINED.
DR EMBL; U79657; AAB40722.1; JOINED.
DR EMBL; U79658; AAB40722.1; JOINED.
SQ SEQUENCE 1411 AA; 144311 MW; 2DB213FA CRC32;

Query Match 38.4%; Score 63; DB 4; Length 1411;
Best Local Similarity 41.7%; Pred. No. 6.76e+00;
Matches 10; Conservative 7; Mismatches 6; Indels 1; Gaps 1;

Db 1170 AKPO-QAAGMLSPKTKGKEASGT 1192
QY 317 AKTEASPRMLDPQTEKEASGG 340

RESULT 9
ID O01385 PRELIMINARY; PRT; 2346 AA.
AC O01385;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE TPR HOMOLOG.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RA ZIMOWSKA G., ARIS J.P., PADDY M.R.;
RL J. Cell Sci. 100:0-0(0).
DR EMBL; U91980; AAC47506.1; -.
DR HSSP; P02632; ICB1.
DR FLYBASE; FBgn0013756; Bx34.
SQ SEQUENCE 2346 AA; 262363 MW; AC4FA447 CRC32;

```

Query Match 37.8%; Score 62; DB 5; Length 2346;
 Best Local Similarity 32.0%; Pred. No. 9.86e+00;
 Matches 8; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

Db 1121 RSNAEAKLLDKERESEKRVSDLH 1145
 QY 318 KTEASPRMLDPQTEKEASGSLQ 342

RESULT 10
 ID O43508 PRELIMINARY; PRT; 249 AA.
 AC O43508;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE TWEAK.
 GN APO3L.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-TONSIL;
 RX MEDLINE; 98070415.
 RA CHICHEPORTICHE Y., BOURDON P.R., XU H., HSU Y.-M., SCOTT H.,
 RA HESION C., GARCIA I., BROWNING J.L.;
 RT "TWEAK, a new secreted ligand in the tumor necrosis factor family that
 weakly induces apoptosis."
 RL J. Biol. Chem. 273:32401-32410(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 9828355.
 RA MARSTERS S.A., SHERIDAN J.P., PITTI R.M., BRUSH J., GODDARD A.,
 RA ASHKENAZI A.;
 RT "Identification of a ligand for the death-domain-containing receptor
 Ap03."
 RL Curr. Biol. 8:525-528(1998).
 DR EMBL; AF030099; AAC51923.1; -.
 DR EMBL; AF035872; AAC39724.1; -.
 SQ SEQUENCE 249 AA; 27216 MW; 3AEAC1F1 CRC32;

Query Match 37.2%; Score 61; DB 4; Length 249;
 Best Local Similarity 40.9%; Pred. No. 1.43e+01;
 Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Db 64 EQQPSSELPNPTESQDPAPFL 85
 QY 320 EAESPRMLDPQTEKEASGSLQ 341

RESULT 11
 ID O82805 PRELIMINARY; PRT; 371 AA.
 AC O82805;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE KNOTTED I CLASS HOMODOMAIN PROTEIN.
 GN KN16 OR HOP1.
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
 OC Pisum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA GOURLAY C.W., HOFER J., ELLIS N.;
 RT "Cloning and characterisation of two knotted 1 class homeodomain
 protein cDNAs from pea."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-SHOOT TIPS;

GILES J.E., VILLANI P.J., DEMASON D.A.;
 RT "A class 1 Knox full-length cDNA from pea shoot tips (Accession No.
 RT AF063307) (PGR98-115).";
 RL Plant Physiol. 117:1125-1125(1998).
 DR EMBL; AF080104; AAC33008.1; -.
 DR EMBL; AF063307; AAC32262.1; -.
 DR MENDEL; 33368; Pissa; 1169; 33368.
 DR PFAM; PF00046; homeobox; 1.
 KW Homeobox; DNA-binding; Nuclear protein.
 SQ SEQUENCE 371 AA; 41854 MW; D607DCE2 CRC32;

Query Match 37.2%; Score 61; DB 10; Length 371;
 Best Local Similarity 47.1%; Pred. No. 1.43e+01;
 Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 244 MIDPOAEDRELKQQLR 260
 QY 326 MLDPOTEKEASGSLQ 342

RESULT 12
 ID P79948 PRELIMINARY; PRT; 375 AA.
 AC P79948;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE LUNATIC FRINGE.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae;
 OC Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96279654.
 RA WU J.-Y., WEN L., ZHANG W.-J., RAO Y.;
 RT "The secreted product of Xenopus gene lunatic fringe, a vertebrate
 signaling molecule."
 RL Science 273:355-358(1996).
 DR EMBL; U77640; BAB19225.1; -.
 SQ SEQUENCE 375 AA; 42133 MW; 575966D0 CRC32;

Query Match 37.2%; Score 61; DB 13; Length 375;
 Best Local Similarity 21.7%; Pred. No. 1.43e+01;
 Matches 5; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

Db 27 DQSRHMLETQSDHEPGSAAVH 49
 QY 320 EAESPRMLDPQTEKEASGSLQ 342

RESULT 13
 ID O9WRT9 PRELIMINARY; PRT; 829 AA.
 AC O9WRT9;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE GLYCOPROTEIN B.
 OS Macaca mulatta rhadinovirus 17577.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 99174001.
 RA SEARLES R.P., BERGQUAM E.P., AXTHELM M.K., WONG S.W.;
 RT "Sequence and genomic analysis of a Rhesus macaque rhadinovirus with
 similarity to Kaposi's sarcoma-associated herpesvirus/human
 herpesvirus 8."
 RL J. Virol. 73:3040-3053(1999).
 DR EMBL; AF083501; AAD21335.1; -.
 SQ SEQUENCE 829 AA; 93300 MW; 89F488CD CRC32;

Query Match 37.2%; Score 61; DB 14; Length 829;
 Best Local Similarity 29.6%; Pred. No. 1.43e+01;

Matches 8; Conservative 13; Mismatches 5; Indels 1; Gaps 1;

Db 741 TNAIAQAPIRMYPIDIDKMQPSGGKVD 767

QY 317 AKTEAESPRMLDPQTEKEASGGSLQ 342

RESULT 14

ID O39233 PRELIMINARY; PRT; 829 AA.

AC O39233; 01-JAN-1998 (TRENBLrel. 05, Created)

DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)

DI 01-JAN-1998 (TRENBLrel. 08, Last annotation update)

DE GLYCOPROTEIN B.

OS rhesus monkey rhadinovirus H26-95.

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Gammaherpesvirinae; Rhadinovirus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-H26-95;

RX MEDLINE; 98037693.

RA DESROSIERS R.C., SASSEVILLE V.G., CZAJAK S.C., ZHANG X.,

RA MANSFIELD K.G., KAUR A., JOHNSON R.P., LACKNER A.A., JUNG J.U.,

RT "A herpesvirus of rhesus monkeys related to the human Kaposi's

RT sarcoma-associated herpesvirus.";

RL J. Virol. 71:9764-9769(1997).

DR EMBL; AF029302; AAC58686.1; -

DR PFAM; PF00606; Glycoprotein B; 1.

SQ SEQUENCE 829 AA; 93250 MW; C162B94C CRC32;

Query Match 37.2%; Score 61; DB 14; Length 829;

Best Local Similarity 29.6%; Pred. No. 1.43e+01;

Matches 8; Conservative 13; Mismatches 5; Indels 1; Gaps 1;

Db 741 TNAIAQAPIRMYPIDIDKMQPSGGKVD 767

QY 317 AKTEAESPRMLDPQTEKEASGGSLQ 342

RESULT 15

ID O92421 PRELIMINARY; PRT; 236 AA.

AC O92421;

DT 01-MAY-1999 (TRENBLrel. 10, Created)

DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)

DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)

DE HYPOTHETICAL 24.0 KD PROTEIN.

GN SC29.09C.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RA SAUNDERS D.C., HARRIS D.;

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RA BENTLEY S.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RX MEDLINE; 97000351.

RA REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,

RA KINASHI H., HOPWOOD D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for

RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).

DR EMBL; AL035707; CAB38884.1; -

DR Hypothetical protein.

SQ SEQUENCE 236 AA; 23972 MW; 5AE00823 CRC32;

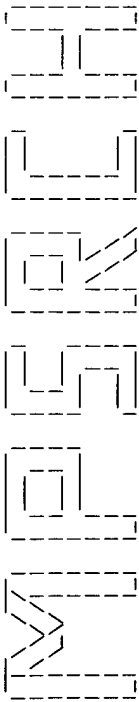
Query Match 36.6%; Score 60; DB 2; Length 236;
Best Local Similarity 36.0%; Pred. No. 2.07e+01;
Matches 9; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

Db 193 ARADAELDRMADFADFAERRAAGRAL 217

QY 317 AKTEAESPRMLDPQTEKEASGGSL 341

Search completed: Wed May 10 14:13:10 2000

Job time : 238 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 14:22:14 2000; MasPar time 2.78 Seconds
Tabular output not generated. 59.542 Million cell updates/sec

Title: >US-09-376-430-2
Description: (347-353) from US09376430A.pep (24 of 25)
Perfect Score: 50
Sequence: 1 PLOGGDV 7

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:geneseq

Statistics: Mean 14.565; Variance 35.951; scale 0.405

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	43	86.0	427	1 W26786	Streptococcus pneumoni	8.28e+01
2	43	86.0	436	1 W32287	Streptococcus pneumoni	8.28e+01
3	43	86.0	447	1 W24487	Class II EPSPs for gly	8.28e+01
4	43	86.0	447	1 W34688	Class II EPSP synthase	8.28e+01
5	43	86.0	447	1 W17619	Synechocystis sp. PC6	8.28e+01
6	42	84.0	639	1 W26530	Trypanosoma cruzi anti	1.11e+02
7	40	80.0	408	1 W89442	Mycoplasma hominis arg	2.01e+02
8	39	78.0	57	1 W34227	SH3 domain A of mouse	2.68e+02
9	39	78.0	299	1 W34410	Mouse Met proto-oncoge	2.68e+02
10	39	78.0	299	1 R50090	Sequence of mouse Met	2.68e+02
11	39	78.0	367	1 W60075	Saccharomyces cerevisi	2.68e+02
12	39	78.0	403	1 W68561	S. peucetius dnrX gene	2.68e+02
13	39	78.0	454	1 W98730	H. pylori GHPO 906 pro	2.68e+02
14	39	78.0	459	1 Y11069	H. pylori ORF 12a6162	2.68e+02
15	39	78.0	686	1 P3711	Sequence of pseudorabi	2.68e+02
16	39	78.0	788	1 W05393	Mouse SH3P12 protein.	3.58e+02
17	38	76.0	249	1 P91215	Domain VI of mouse lam	3.58e+02
18	38	76.0	421	1 W86326	Kidney injury associat	3.58e+02
19	38	76.0	760	1 W29490	Programmed cell death	3.58e+02
20	38	76.0	1607	1 W50897	Mouse laminin G1 chain	3.58e+02
21	38	76.0	1609	1 W50898	Human laminin G1 chain	3.58e+02
22	37	74.0	34	1 R43826	Anti-GAT antibody immu	4.75e+02
23	37	74.0	46	1 Y13195	Human secreted protein	4.75e+02

24	37	74.0	106	1 W74752	Human secreted protein	4.75e+02
25	37	74.0	131	1 Y12477	Human 5' EST secreted	4.75e+02
26	37	74.0	235	1 W60078	Drosophila melanogaste	4.75e+02
27	37	74.0	249	1 W84049	Rice OsMADS7 protein i	4.75e+02
28	37	74.0	398	1 R55706	Galactosyltransferase.	4.75e+02
29	37	74.0	398	1 R28838	HeLa cell galactosyltr	4.75e+02
30	37	74.0	536	1 P70142	Sequence of a region o	4.75e+02
31	37	74.0	537	1 R44943	Avian infectious bronc	4.75e+02
32	37	74.0	601	1 R33747	XR79.	4.75e+02
33	37	74.0	767	1 R55708	Glycosyltransferase hy	4.75e+02
34	37	74.0	767	1 R55709	Glycosyltransferase hy	4.75e+02
35	37	74.0	816	1 R85670	WD-40 domain-contg. Mu	4.75e+02
36	37	74.0	1093	1 R40815	Truncated IBV spike pr	4.75e+02
37	37	74.0	1162	1 P60720	Sequence of the S1 sig	4.75e+02
38	37	74.0	1162	1 P70137	Sequence of a region o	4.75e+02
39	37	74.0	1162	1 W06829	Major neutralising ant	4.75e+02
40	37	74.0	1362	1 W99481	Murine P/CIP protein.	4.75e+02
41	37	74.0	1447	1 W81029	Murine PCIP protein (o	4.75e+02
42	37	74.0	1471	1 W74585	Calcium independent re	4.75e+02
43	37	74.0	1481	1 W95160	Amino acid sequence of	4.75e+02
44	37	74.0	1481	1 W89299	Human rchd528 protein.	4.75e+02
45	37	74.0	3567	1 R44431	eryA region polypeptid	4.75e+02

ALIGNMENTS

RESULT 1
ID W26786 standard; Protein; 427 AA.
AC W26786;
DT 22-JUN-1998 (first entry)
DE Streptococcus pneumoniae Mur A-1.
KW Mur A-1; UDP-N-acetylglucosamine enolpyruvyl-transferase;
KW otitis media; conjunctivitis; pneumonia; bacteraemia; meningitis;
KW sinusitis; pleural empyema; endocarditis; genetic immunisation;
KW infection; bacteriostatic; bacteriocide; diagnosis; therapy;
KW vaccine.
OS Streptococcus pneumoniae.
PN EP-816502-A2.
PD 07-JAN-1998.
PF 02-JUL-1997; 304806.
PR 21-OCT-1996; US-028976.
PR 03-JUL-1996; GB-013907.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI (SMIK) SMITHKLINE BEECHAM PLC.
PI Wallis NG.
DR WPI: 98-054915/06.
DR N-FSDB; T12034.
PT Mur A-1 polypeptide from Streptococcus pneumoniae - useful for e.g.
PT treating patients requiring Mur A-1, diagnosis of e.g. pneumonia and
PT meningitis and screening of inhibitory compounds
PS Claim 13: Page 22: 28pp: English.
CC This UDP-N-acetylglucosamine enolpyruvyl-transferase, designated
CC Mur A-1, is a novel enzyme of Streptococcus pneumoniae. Its amino
CC acid sequence was deduced from an isolated polynucleotide (see
CC T12034). Mur A-1 polypeptides can be administered therapeutically,
CC especially by expressing encoding DNA in vivo, to treat patients
CC requiring Mur A-1 (claimed). They are useful e.g. to treat
CC infections, especially bacterial infections. They can be used
CC to identify compounds which bind to and inhibit Mur A-1 activity
CC (claimed), e.g. potential new antibiotics, or compounds enhancing
CC Mur A-1 activity. Antagonists can be administered to treat
CC bacterial infections. Diagnostic assays for levels of Mur A-1
CC polypeptides in a host sample can be undertaken to detect diseases
CC related to Mur A-1 expression e.g. otitis media, conjunctivitis,
CC pneumoniae, bacteraemia, sinusitis, pleural empyema, endocarditis,
CC and especially meningitis, or to diagnose sequence mutations.
CC They are also for inducing an immunological response that protects
CC an animal from such disease (claimed).
SQ Sequence 427 AA.

Query Match 86.0%; Score 43; DB 1; Length 427;
Best Local Similarity 71.4%; Pred. No. 8.28e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

Db 363 PLOGAEV 369
QY 347 PLOGGDV 353

RESULT 2
ID W32287 standard; Protein; 436 AA.
AC W32287;
DE 20-MAR-1998 (first entry)
DE Streptococcus pneumoniae Mura protein.
KW Peptidoglycan biosynthetic Mura; stem peptide biosynthetic gene;
KW inhibitor; substrate.
OS Streptococcus pneumoniae.
PN US5691161-A.
PD 25-NOV-1997.
PF 01-AUG-1996; 691129.
PR 01-AUG-1996; US-691129.
PA (ELIL ) LILLY & CO ELI.
PI Peery RB, Skatrud PL;
DR WPI; 98-017632/02.
DR N-PSDB; T86547.
PT Streptococcus pneumoniae Mura protein - useful in screening assay
PT for Mura inhibitors
PS Claim 1; Columns 15-18; 12pp; English.
CC The present sequence represents a novel Streptococcus pneumoniae Mura
CC protein. It can be used in a new method to identify compounds that
CC inhibit stem peptide biosynthesis (Mura inhibitors) comprising mixing
CC Mura with a substrate and a test compound in a buffer, measuring the
CC amount of a product formed and comparing the amount with that formed in
CC the absence of the test compound.
SQ Sequence 436 AA;

Query Match 86.0%; Score 43; DB 1; Length 436;
Best Local Similarity 71.4%; Pred. No. 8.28e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 372 PLOGAEV 378
QY 347 PLOGGDV 353

RESULT 3
ID W24487 standard; Protein; 447 AA.
AC W24487;
DE 02-OCT-1997 (first entry)
DE Class II EPSPS for glyphosate resistant plant production.
KW 5-enolpyruvylshikimate-3-phosphate synthase; EPSPS; Class II;
KW glyphosate resistant; transgenic plant; herbicide; shikimic acid;
KW fusion protein; soybean; corn; wheat; barley; cotton; sugarbeet;
KW tobacco.
OS Synecocystis sp.; PCC6803.
PN US5633435-A.
PD 27-MAY-1997.
PF 31-AUG-1990; 576537.
PR 31-AUG-1990; US-576537.
PR 28-AUG-1991; US-749611.
PR 13-SEP-1994; US-306063.
PA (MONS ) MONSANTO CO.
PI Barry GF, Kishore GM, Padgett SR, Stallings WC;
DR WPI; 97-297418/27.
DR N-PSDB; T77328.
PT New isolated 5-enolpyruvylshikimate-3-phosphate synthase gene -
PT used for transforming plants to produce plants which are tolerant to
PT glyphosate herbicide
PS Disclosure: Column 151-154; 154pp; English.
CC W24487 shows the sequence of a Class II 5-enolpyruvylshikimate-3-
CC phosphate synthase (EPSPS) enzyme. Class II EPSPS enzymes are tolerant
CC to glyphosate herbicides. EPSPS and sequences encoding it are used for
CC the production of herbicide resistant (glyphosate-tolerant) plants
CC such as corn, wheat, rice, barley, soybean, cotton, sugarbeet, oilseed
CC rape, canola, flax, sunflower, potato, tobacco, tomato, alfalfa, poplar,
CC pine, eucalyptus, apple, lettuce, peas, lentils, grape and turf grasses.

SQ Sequence 447 AA;

Query Match 86.0%; Score 43; DB 1; Length 447;
Best Local Similarity 71.4%; Pred. No. 8.28e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 393 PLOGAEV 399
QY 347 PLOGGDV 353

RESULT 4
ID W34688 standard; Protein; 447 AA.
AC W34688;
DE 17-FEB-1998 (first entry)
DE Class II EPSP synthase (EPSPS) from Synecocystis sp. PCC6803.
KW 5-enolpyruvylshikimate-3-phosphate synthase; EPSPS; glyphosate;
KW shikimic acid pathway; phosphoenolpyruvic acid; 3-phosphoshikimic acid;
KW 5-enolpyruvyl-3-phosphoshikimic acid; herbicide; transgenic plant;
KW glyphosate resistance gene; glyphosate-tolerance; promoter.
OS Synecocystis sp. PCC6803.
PN US5627061-A.
PD 06-MAY-1997.
PF 07-JUN-1995; 476008.
PR 13-SEP-1994; US-306063.
PR 31-AUG-1990; US-576537.
PR 28-AUG-1991; US-749611.
PA (MONS ) MONSANTO CO.
PI Barry GF, Kishore GM, Padgett SR, Stallings WC;
DR WPI; 97-271315/24.
DR N-PSDB; T93795.
PT Production of glyphosate-herbicide tolerant plants - using DNA
PT encoding class II 5-enolpyruvylshikimate-3-phosphate synthase
PT enzyme(s)
PS Disclosure; Fig 21; 151pp; English.
CC W34688-89 represent a new class of glyphosate synthases (EPSPS). These novel
CC 5-enolpyruvylshikimate-3-phosphate synthases (EPSPS). Class I EPSPS enzymes, and
CC EPSPS enzymes have little homology with known Class I EPSPS enzymes, and
CC belong to a new class, Class II. The present sequence was isolated from
CC Synecocystis sp. PCC6803. The EPSPS enzymes are part of the shikimic
CC acid pathway, which leads to the biosynthesis of aromatic compounds.
CC EPSPS converts phosphoenolpyruvic acid (PEP) and 3-phosphoshikimic acid
CC to 5-enolpyruvyl-3-phosphoshikimic acid, and is inhibited by the
CC herbicide glyphosate. It would be useful to produce transgenic crops
CC containing glyphosate resistance genes so that glyphosate-containing
CC herbicides can be applied to selectively kill weeds. The novel EPSPS
CC enzymes exhibit a low Km for PEP and a high Ki for glyphosate, such that
CC when introduced into a plant, the plant is made glyphosate-tolerant, and
CC EPSPS enzyme activity is not affected. These class II EPSPS enzymes are
CC fused to a chloroplast transit peptide to target the protein into the
CC chloroplast, which is the site for the shikimic acid pathway. In
CC addition, the EPSPS gene is cloned into a plant under the control of a
CC promoter such as figwort mosaic virus promoter or the cauliflower mosaic
CC virus promoter, so that expression is enhanced.
SQ Sequence 447 AA;

Query Match 86.0%; Score 43; DB 1; Length 447;
Best Local Similarity 71.4%; Pred. No. 8.28e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 393 PLOGAEV 399
QY 347 PLOGGDV 353

RESULT 5
ID W71619 standard; Protein; 447 AA.
AC W71619.
DT 14-DEC-1998 (first entry)
DE Trypanosoma sp. PC6803 Class II EPSPS.
KW 5-enolpyruvylshikimate-3-phosphate synthase; class II EPSPS enzyme;
KW glycosylase herbicide; transformed bacteria; class I EPSPS enzyme;
KW resistance; inhibition; 5-enolpyruvyl-3-phosphoshikimic acid;
KW N-phosphonomethylglycine.
OS Trypanosoma sp.
PN US5804425-A.
PD 08-SEP-1998. 833485.
PF 07-APR-1997; US-306063.
PR 13-SEP-1994; US-306063.
PR 31-AUG-1990; US-576537.
PR 28-AUG-1991; US-749611.
PR 07-APR-1997; US-833485.
PA (MONS ) MONSANTO CO.
PI Barry GF, Kishore GM, Padgett SR, Stallings WC;
WPI: 98-505657/43.
DR N-PSDB: V58020.
PT Glycosylase resistant 5-enolpyruvylshikimate-3-phosphate synthase -
PT useful for characterisation of the enzyme to determine inhibition
PT data values
PS Disclosure: Fig 21; 152pp; English.
CC An isolated 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) has the
CC sequence domains (I), (II), (III) and (IV): -R-X1-H-X2-E- (I),
CC -S-A-Q-X4-K- (III); and -N-X5-T-R- (IV). Where X1 = G,
CC G-D-K-X3- (II); -S-A-Q-X4-K- (III); and -N-X5-T-R- (IV). Where X1 = G,
CC S, T, C, Y, N, Q, D or E; X2 and X3 = S or T; X4 and X5 = A, R, N, D, C,
CC E, G, H, I, L, K, M, F, P, S, T, W, Y or V. The EPSPS enzyme,
CC produced by recombinant methods, can be used in kinetic studies to
CC determine Ki and Km values of the enzyme for its characterisation. The
CC enzyme is normally used for the production of 5-enolpyruvyl-3-
CC phosphoshikimic acid in plants, and most forms of the enzyme are
CC inhibited by N-phosphonomethylglycine ('glyphosate') herbicides.
CC Inhibition data enables more accurate values of concentrations of
CC herbicide to be used when growing the plant without being detrimental to
CC it. This enables the plant to be grown in the presence of the herbicide,
CC being used to inhibit the growth of undesired plants. The present
CC sequence represents a Class II EPSPS from bacterial isolate
CC Trypanosoma sp. PC6803.
CC Synchocystis sp.
SQ Sequence 447 AA;

Query Match 86.0%; Score 43; DB 1; Length 447;
Best Local Similarity 71.4%; Pred. No. 8.28e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 393 PLOGAEV 399
QY 347 PLOGGDV 353

RESULT 6
ID W26530 standard; Protein; 639 AA.
AC W26530.
DT 12-JAN-1998 (first entry)
DE Trypanosoma cruzi antigen TcH12.
KW Antigen; epitope; vaccine; protective immunity; Chagas disease;
KW diagnosis; therapy; immunosassay.
OS Trypanosoma cruzi Tulahuan strain C2.
FH Key Location/Qualifiers
FT Misc-difference 236
FT /note= "unidentified amino acid residue"
FT Misc-difference 536
FT /note= "unidentified amino acid residue"
FT Misc-difference 539
FT /note= "unidentified amino acid residue"
FT Misc-difference 596
FT /note= "unidentified amino acid residue"

```

```

FT W09718475-A1.
PN 22-MAY-1997.
PF 14-NOV-1996; U18624.
PR 14-NOV-1995; US-557309.
PA (CORI-) CORIXA CORP.
PI Houghton RL, Lodes MJ, Reed SG, Skeiky YAW;
WPI: 97-289413/26.
DR N-PSDB: T69157.
PT Diagnosing Trypanosoma cruzi infection by detecting antibodies to
PT novel antigens - which are useful in vaccines to provide protective
PT immunity against Chagas' disease
PS Disclosure: Page 60-63; 110pp; English.
CC This polypeptide sequence comprises Trypanosoma cruzi full-length
CC antigen TcH12, identified by sequencing a DNA clone (see T69157)
CC obtained by screening a Trypanosoma cruzi genomic expression library
CC with pools of sera from infected individuals. T. cruzi antigens
CC (see W26530-41), or epitope-containing repeat sequences (see W15094-
CC 102, W19079-86 and W26542-44) of native antigens, can be used in a
CC variety of immunoassays for detecting T. cruzi infection in a
CC blood, serum, plasma, saliva, cerebrospinal fluid or urine sample.
CC The polypeptides are also useful in vaccines and pharmaceutical
CC compositions for inducing protective immunity against Chagas
CC disease. They can be produced by expression in transformed or
CC transfected host cells.
SQ Sequence 639 AA;

Query Match 84.0%; Score 42; DB 1; Length 639;
Best Local Similarity 85.7%; Pred. No. 1.11e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 415 PLAGGDV 421
QY 347 PLOGGDV 353

RESULT 7
ID W89442 standard; protein; 408 AA.
AC W89442.
DT 18-MAR-1999 (first entry)
DE Mycoplasma hominis arginine deiminase.
KW Arginine deiminase; Mycoplasma arginini; Mycoplasma arthritis;
KW Mycoplasma hominis; modified; polyethylene glycol; cancer; metastasis;
KW inhibition; melanoma; hepatoma; sarcoma; enhanced circulating half life.
OS Mycoplasma hominis.
PN W09851784-A1.
PD 19-NOV-1998.
PF 12-MAY-1998; U09575.
PR 13-FEB-1998; US-023809.
PR 12-MAY-1997; US-046200.
PA (PHOB-) PHOENIX PHARMACOLOGICS INC.
PI Clark MA;
WPI: 99-045227/04.
PT New compound comprising arginine deiminase - covalently bonded via
PT linking group to polyethylene glycol, to enhance the half life of
PT arginine by this modification
PS Claim 6; Fig 1; 30pp; English.
CC The present invention describes: (1) a compound comprising arginine
CC deiminase (AD) covalently bonded via linking group to polyethylene
CC glycol (PEG), and having a molecular weight 12-40 kDa; and (2) a
CC composition as above, but where the linking group is selected from a
CC malimide group, an amide group, an imide group, a carbamate group, an
CC ester group, an epoxy group, a carboxyl group, a hydroxyl group, a
CC carbohydrate, a tyrosine group, a cysteine group and/or a histidine
CC group. AD can be used in the treatment of tumours, e.g. melanomas,
CC hepatomas and sarcomas, and to inhibit metastasis. The modified AD has
CC an enhanced circulating half life.
SQ Sequence 408 AA;

Query Match 80.0%; Score 40; DB 1; Length 408;
Best Local Similarity 71.4%; Pred. No. 2.01e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

Db 210 PLEGDV 216
 QY 347 PLOGGDV 353

RESULT 8
 ID W34227 standard; Protein; 57 AA.

AC W34227;
 DE SH3 domain A of mouse SH3P12 protein.
 KW src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
 KW cellular signalling element; cellular structural element; malignancy;
 KW protein identification; functional domain; protein screening;
 KW cellular signal transduction process.
 OS Mus musculus.
 PN W09631625-A1.
 PD 10-OCT-1996.
 PF 04-APR-1996; U04454.
 PR 03-APR-1996; US-630915.
 PR 07-APR-1995; US-417872.
 PA (CYTO-) CYTOGEN CORP.
 PA (UYN-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
 DR WPI: 96-465045/46.
 PT Identifying polypeptide(s) having specific functional domain (esp.
 PT SH3 domain). comprises detecting selective binding to recognition
 PT unit, regardless of sequence homology
 PS Claim 55; Fig 10a; 174pp; English.
 CC W34220-W34253 represent Src-homology region 3 (SH3) domains from human
 CC and mouse proteins, that can be used in the method of the invention. This
 CC sequence represents a SH3 domain of the mouse SH3P12 protein (see W03393
 CC for full length protein). SH3 domain containing proteins play a role in
 CC signalling and structural elements of cells. The method of the invention
 CC is for identifying polypeptides containing functional domains of interest
 CC (especially SH3 domains). It comprises contacting a multivalent
 CC recognition unit (RU) complex with a number of peptides and identifying
 CC polypeptides having a selective binding affinity for the RU complex. The
 CC method is based on functional similarities and does not rely on sequence
 CC similarities. Prior methods only gave limited success for identifying
 CC proteins containing an SH3 domain due to the minimal sequence homology
 CC among known SH3 proteins. Multivalent RU complexes are particularly
 CC suited to screening for polypeptides containing functional domains that
 CC are similar to, but not identical in sequence to, the original target
 CC functional domain. The new method enables proteins having a common
 CC function to be identified. Identification of novel SH3 proteins will be
 CC useful for a better understanding of cell growth, malignancy, signal
 CC transduction processes, etc. New candidate drugs can be identified, and
 CC their specificities (e.g. pharmacological activities) can be assessed
 CC using the method of the invention.
 SQ Sequence 57 AA;

Query Match 78.0%; Score 39; DB 1; Length 57;
 Best Local Similarity 85.7%; Pred. No. 2.68e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 20 PLOGGDV 26
 QY 347 PLOGGDV 353

RESULT 9

ID W94410 standard; protein; 299 AA.

AC W94410;
 DE Mouse Met proto-oncogene protein 501-850.
 KW Mouse; Met; c-met; proto-oncogene; hepatocyte growth factor; HGF;
 KW scatter factor; SF; HGF/SF; tumorigenic; tumour cell metastasis;
 KW tyrosine kinase receptor; paracrine signalling system; autocrine;
 KW sarcomagenesis; sarcoma; human.
 OS Mus sp.
 PN US5871959-A.
 PD 16-FEB-1999.

PF 18-NOV-1994; 344710.
 PR 18-NOV-1994; US-344710.
 PR 27-DEC-1989; US-457556.
 PR 14-SEP-1990; US-582063.
 PR 18-JAN-1991; US-642971.
 PR 18-SEP-1992; US-946061.
 PR 20-MAY-1994; US-246667.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Faletto DL, Oskarsson M, Rong S, Tsarfaty I, Woude GV;
 DR WPI: 99-186642/14.
 PT Production of human hepatocyte growth factor or scatter factor
 PT comprises transfecting cells with DNA encoding the factors and
 PT growing the cells in mice
 PS Disclosure: Fig 6B; 51pp; English.
 CC A method has been developed for producing human hepatocyte growth
 CC factor/scatter factor (HGF/SF). The method comprises: (a) transfecting
 CC mammalian cells, especially NIH/3T3 cells with DNA encoding HGF/SF and
 CC Met operatively linked to a control sequence; (b) introducing the
 CC transfected cells into an immunocompromised mouse to generate a primary
 CC tumour; (c) transplanting and propagating cells from the primary tumour in
 CC vitro; (d) selecting cells which express HGF/SF and Met, where Met is
 CC expressed at higher levels than HGF/SF; (e) introducing the selected
 CC cells into a mouse to provide a secondary tumour; (f) isolating HGF/SF
 CC propagating the secondary tumour cells in vitro; and (g) isolating HGF/SF
 CC produced by the cells. The present sequence represents a mouse Met
 CC protein from the present invention. Studies have revealed that most
 CC human tumour and tumour cell lines of mesenchymal origin inappropriately
 CC express Met and/or HGF, suggesting that Met plays a role in
 CC sarcomagenesis and that both autocrine and paracrine modes of
 CC stimulation can occur in human primary tumours. This indicates that
 CC Met-HGF/SF autocrine and/or paracrine signalling contributes to the
 CC tumorigenic process in human tumours of mesenchymal origin, such as
 CC sarcomas.
 SQ Sequence 299 AA;

Query Match 78.0%; Score 39; DB 1; Length 299;
 Best Local Similarity 71.4%; Pred. No. 2.68e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 74 PLEGTV 80
 QY 347 PLOGGDV 353

RESULT 10
 ID R50090 standard; peptide; 299 AA.
 AC R50090;
 DT 03-OCT-1994 (first entry)
 DE Sequence of mouse Met between the NdeI - PvuII sites, residues
 DE 500-849.
 KW Hepatocyte growth factor; HGF; scatter factor; SF; Met; metastasis;
 KW tumour cell
 OS Mus musculus.
 PN W09406909-A.
 PD 31-MAR-1994.
 PF 15-SEP-1993; U08531.
 PF 18-SEP-1992; US-946061.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Faletto DL, Oskarsson M, Rong S, Tsarfaty I, Woude GV;
 DR WPI: 94-118458/14.
 PT Preventing tumour cell metastasis - using a substance which
 PT prevents hepatocyte growth factor-scatter factor from binding
 PT with Met
 PS Disclosure: Fig 6B; 42pp; English.
 CC The computer-generated predicted structure of the AA sequence
 CC between the conserved NdeI-PvuII sites in the human Met and the
 CC mouse Met were compared. The two regions where the AA sequence is
 CC less conserved were highlighted. Unfortunately, this rendered them
 CC illegible and the AAs in these two regions are denoted by X in the
 CC sequence. The less conserved domains within the NdeI-PvuII segment of
 CC cDNA conferring transforming potential onto Met either reflect
 CC domains directly involved in ligand binding, or through structural
 CC characteristics modulate either ligand binding or activation of the

CC receptor following ligand binding.
SQ Sequence 299 AA;

Query Match 78.0%; Score 39; DB 1; Length 299;
Best Local Similarity 71.4%; Pred. No. 2.68e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 74 PLEGGTV 80
||:|:|
QY 347 PLOGGDV 353

RESULT 11
ID W60075 standard; Protein; 367 AA.

AC W60075;
DE Saccharomyces cerevisiae RNA 3'-terminal phosphate cyclase.
KW RNA 3'-terminal cyclase; RNA metabolism; synthesis; study; ds.
OS Saccharomyces cerevisiae
FH Key Location/Qualifiers
FT Peptide 1..367 /note= "mature peptide"
FT WO9745535-A1.
PN PD 04-DEC-1997.
PF 20-MAY-1997; E02566.
PR 24-MAY-1996; US-018335.
PA (NOVS) NOVARTIS AG.
PI Filipowicz W, Genschik P;
DR WPI; 98-286352/25.
DR N-PSDB; V34545.
PT New isolated RNA 3'-terminal phosphate cyclase nucleic acids - used
PT to develop products for the study of RNA synthesis and metabolism in
PT cells and for detection, diagnosis and therapy
PS Disclosure; Page 42-43; 71pp; English.
CC The sequence is that of RNA 3'-terminal phosphate cyclase
CC (RNA 3'-TPC). RNA 3'-TPC catalyses the conversion of 3' phosphate
CC to a 2',3'-phosphodiester at the 3' end of RNA molecules. This
CC enables the study of RNA synthesis and metabolism in the cell, as
CC well as in vitro RNA manipulation. RNA 3'-TPC can also be used
CC to develop products for detection, diagnosis and therapy of
CC diseases associated with RNA 3'-TPC.
SQ Sequence 367 AA;

Query Match 78.0%; Score 39; DB 1; Length 367;
Best Local Similarity 71.4%; Pred. No. 2.68e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 160 PLOGGEV 166
||:|:|
QY 347 PLOGGDV 353

RESULT 12
ID W68561 standard; Protein; 403 AA.

AC W68561;
DE S. peucetius dnrX gene product.
KW dnrX; dnrX; recombinant; microorganism; doxorubicin; daunorubicin;
KW metabolism; baumycin; secretion; anti-tumour compound.
OS Streptococcus peucetius.
PN WO9839458-A1.
PD 11-SEP-1998.
PF 05-MAR-1998; US-901306.
PR 28-JUL-1997; US-812412.
PR 06-MAR-1997; US-812412.
PA (PHAA) PHARMACIA & UPJOHN SPA.
PI Colombo AL, Filippini S, Fonstein L, Hutcheson RC,
PI Lomovskaya N;
DR WPI; 98-506367/43.
DR N-PSDB; V34002.
PT Production of doxorubicin and daunorubicin - using a recombinant.
PT daunorubicin producing microorganism in which at least one dnrX or
PT dnrX metabolism gene has been inactivated

PS Disclosure; Page 29-32; 48pp; English.
CC This sequence represents the dnrX gene product from the microorganism
CC Streptomyces peucetius. The invention relates to a recombinant
CC microorganism which produces doxorubicin (DoR), in which at least 1
CC daunorubicin (DaR) metabolism gene selected from dnrU and dnrX, has been
CC inactivated. The microorganism is preferably S. peucetius WHH1658,
CC WHH1654 and WHH1662. The function of the DaR gene is inactivated by
CC insertion of a gene (preferably neomycin/kanamycin resistance gene aphII)
CC into the daunorubicin metabolism gene. The blocking of the function of
CC at least one gene of DaR metabolism can increase DaR and DoR production
CC levels and cause the disappearance of baumycin-like products resulting
CC in DaR and DoR secretion directly into the culture medium. The products
CC can be used in the production of anti-tumour compounds.
SQ Sequence 403 AA;

Query Match 78.0%; Score 39; DB 1; Length 403;
Best Local Similarity 57.1%; Pred. No. 2.68e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 247 PVHGGEV 253
||:|:|
QY 347 PLOGGDV 353

RESULT 13

ID W98730 standard; Protein; 454 AA.
AC W98730;
DE 31-MAR-1999 (first entry)
DE H. pylori GHPO 906 protein.
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
OS Helicobacter pylori.
PN WO9843478-A1.
PD 08-OCT-1998.
PF 01-APR-1998; U06371.
PR 29-JUL-1997; US-902615.
PR 01-APR-1997; US-833457.
PR 24-JUN-1997; US-881227.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR WPI; 98-542293/46.
DR N-PSDB; X14449.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 8; Page 1547-1549; 2054pp; English.
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 454 AA;

Query Match 78.0%; Score 39; DB 1; Length 454;
Best Local Similarity 57.1%; Pred. No. 2.68e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 17 SLHGGEV 23
||:|:|
QY 347 PLOGGDV 353

RESULT 14

ID Y11069 standard; Protein; 459 AA.
AC Y11069;
DE 08-JUN-1999 (first entry)
DE H. pylori ORF 12ae10622.30273255_f1_13 secreted protein.
KW Vaccine; probe; diagnostic; ORF; cell envelope protein;
KW secreted protein; cytoplasmic protein; cellular protein.
OS Helicobacter pylori.

Search completed: Wed May 10 14:22:22 2000
Job time : 8 secs.

PN WO9824475-A1.
PD 11-JUN-1998.
PF 05-DEC-1997; U22104.
PR 14-JUL-1997; US-891928.
PR 05-DEC-1996; US-759625.
PR 25-MAR-1997; US-823745.
PA (ASTR) ASTRA AB.
PI Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;
DR WPI; 98-333051/29.
DR N-PSDB; X30598.
PT New isolated Helicobacter pylori nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of infection by
PT H. pylori and other Helicobacter species
PS Claims 37, 41; Page 234-235; 339pp; English.
CC Recombinant or substantially pure preparations of H. pylori polypeptides
CC are disclosed, together with the nucleic acids encoding them. In all,
CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
CC cytoplasmic proteins, secreted proteins or other cellular proteins.
CC Vaccines containing the nucleic acids or proteins are claimed, as are
CC probes containing at least 8 nucleotides from the nucleic acid
CC sequences. The vaccines are useful for treating or reducing the risk of
CC H. pylori infections, and the probes can be used diagnostically for
CC detecting the presence of Helicobacter in a sample. The products are
CC also of use in screening for compounds having the ability to interfere
CC with the H. pylori life cycle or to inhibit H. pylori infection.
SQ Sequence 459 AA;

Query Match 78.0%; Score 39; DB 1; Length 459;
Best Local Similarity 57.1%; Pred. No. 2.68e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 22 SLHGEV 28

QY 347 PLQGGDV 353

RESULT 15
ID P93711 standard; Protein; 686 AA.
AC P93711.
DT 11-MAR-1992 (first entry)
DE Sequence of pseudorabies virus (PRV) glycoprotein gH.
KW Herpes virus; vaccine; surface polypeptide; Bulbar paralysis;
KW Aujeszky's disease; mad itch; tracheitis; rhinitis.
OS Pseudorabies virus.
PN WO8910965-A.
PD 16-NOV-1989.
PF 18-APR-1989; 901563.
PR 03-MAY-1988; US-189907.
PA (UPJO) UPJOHN CO.
PI Petrovskis EA, Post LE;
DR WPI; 89-356494/48.
DR N-PSDB; N92408.
PT Recombinant DNA - comprises sequence encoding glycoprotein gH of
PT Pseudorabies virus or infectious bovine rhinotracheitis virus
PS Claim 7; Page 26-27 and Chart A, Page 15-17; 31pp; English.
CC The recombinant DNA of the invention (see N92408 and N93701) can be
CC used to diagnose PRV and IBRV infections. The encoded proteins
CC (see P93711 and P93712) or their fragments or derivs., are useful as
CC virus-free vaccine components for protecting animals against PRV or
CC IBRV infections. Host cells transformed with N92408 or N93701
CC operably linked to an expression control sequence are also claimed.
CC The host cell is from a bacterium, fungus plant, or animal cell. Pref.
CC the host is an insect cell infected with a baculovirus.
SQ Sequence 686 AA;

Query Match 78.0%; Score 39; DB 1; Length 686;
Best Local Similarity 83.3%; Pred. No. 2.68e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 614 LRGGDV 619

QY 348 LRGGDV 353



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 14:22:39 2000; MasPar time 45.86 Seconds
1.978 Million cell updates/sec

Tabular output not generated.

Title: >US-09-376-430-2
Description: (347-353) from US09376430A.pep (24 of 25)
Perfect Score: 50
Sequence: 1 PLQGGDV 7

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 13.878; Variance 33.416; scale 0.415

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	43	86.0	436	1	US-08-691- Sequence 2, Applicatio	3.42e+01
2	43	86.0	436	1	US-08-818- Sequence 2, Applicatio	3.42e+01
3	43	86.0	447	1	US-08-476- Sequence 67, Applicati	3.42e+01
4	43	86.0	447	1	US-08-833- Sequence 67, Applicati	3.42e+01
5	43	86.0	447	1	US-08-308- Sequence 67, Applicati	3.42e+01
6	42	84.0	639	2	US-08-557- Sequence 37, Applicatio	4.67e+01
7	40	80.0	410	2	US-09-105- Sequence 8, Applicatio	8.62e+01
8	40	80.0	410	1	US-08-792- Sequence 6, Applicatio	1.17e+02
9	39	78.0	375	1	US-08-255- Sequence 2, Applicatio	1.17e+02
10	39	78.0	403	2	US-08-812- Sequence 2, Applicatio	1.58e+02
11	38	76.0	251	1	US-08-152- Sequence 37, Applicati	1.58e+02
12	38	76.0	252	1	US-08-152- Sequence 38, Applicati	1.58e+02
13	38	76.0	252	2	US-08-125- Sequence 12, Applicati	1.58e+02
14	38	76.0	252	2	US-08-460- Sequence 12, Applicati	1.58e+02
15	38	76.0	760	1	US-08-195- Sequence 2, Applicatio	2.12e+02
16	37	74.0	249	2	US-08-867- Sequence 15, Applicati	2.12e+02
17	37	74.0	398	1	US-08-446- Sequence 2, Applicatio	2.12e+02
18	37	74.0	601	1	US-08-463- Sequence 14, Applicati	2.12e+02
19	37	74.0	601	1	US-08-333- Sequence 14, Applicati	2.12e+02
20	37	74.0	601	1	US-08-694- Sequence 6, Applicati	2.12e+02
21	37	74.0	767	1	US-08-446- Sequence 8, Applicatio	2.12e+02
22	37	74.0	767	1	US-08-446- Sequence 8, Applicatio	2.12e+02
23	37	74.0	816	1	US-08-190- Sequence 54, Applicati	2.12e+02

24	37	74.0	1162	2	US-08-288- Sequence 15, Applicati	2.12e+02
25	37	74.0	1162	2	US-08-023- Sequence 15, Applicati	2.12e+02
26	37	74.0	1162	2	US-08-663- Sequence 15, Applicati	2.12e+02
27	37	74.0	1162	3	PCT-US95-1 Sequence 15, Applicati	2.12e+02
28	37	74.0	1162	2	US-08-362- Sequence 15, Applicati	2.12e+02
29	37	74.0	1481	2	US-08-616- Sequence 40, Applicati	2.12e+02
30	37	74.0	1481	2	US-08-599- Sequence 40, Applicati	2.12e+02
31	37	74.0	3567	2	US-07-642- Sequence 4, Applicatio	2.12e+02
32	36	72.0	228	2	US-08-408- Sequence 21, Applicati	2.85e+02
33	36	72.0	250	2	US-08-125- Sequence 11, Applicati	2.85e+02
34	36	72.0	264	2	US-08-460- Sequence 10, Applicati	2.85e+02
35	36	72.0	589	3	PCT-US91-0 Sequence 6, Applicatio	2.85e+02
36	36	72.0	589	2	US-08-429- Sequence 6, Applicatio	2.85e+02
37	36	72.0	589	2	US-08-431- Sequence 4, Applicatio	2.85e+02
38	36	72.0	616	1	US-08-385- Sequence 2, Applicatio	2.85e+02
39	36	72.0	616	1	US-08-385- Sequence 2, Applicatio	2.85e+02
40	36	72.0	870	1	US-08-785- Sequence 4, Applicatio	2.85e+02
41	36	72.0	1611	2	US-08-804- Sequence 5, Applicatio	2.85e+02
42	36	72.0	3111	2	US-08-125- Sequence 4, Applicatio	2.85e+02
43	36	72.0	3111	2	US-08-125- Sequence 4, Applicatio	2.85e+02
44	36	72.0	3729	2	US-08-804- Sequence 4, Applicatio	2.85e+02
45	36	72.0	4545	2	US-08-804- Sequence 14, Applicati	2.85e+02

ALIGNMENTS

RESULT 1
ID US-08-691-129-2 STANDARD: PRT: 436 AA.
XX
AC xxxxxx
XX
DT XX
DE
DE Sequence 2, Application US/08691129
XX Patent No. 5691161
CC
CC GENERAL INFORMATION:
CC APPLICANT: Skatrud, Paul L.
CC APPLICANT: Peery, Robert B.
CC TITLE OF INVENTION: Peptidoglycan Biosynthetic Gene mura
CC TITLE OF INVENTION: From Streptococcus Pneumoniae
CC NUMBER OF SEQUENCES: 3
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Eli Lilly and Company
CC STREET: Lilly Corporate Center
CC CITY: Indianapolis
CC STATE: Indiana
CC COUNTRY: U.S.
CC ZIP: 46285

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,129
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Webster, Thomas D.
REGISTRATION NUMBER: 39,872
REFERENCE/DOCKET NUMBER: X-9897
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3334
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 436 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 436 AA; 46942 MW; 927638 CN;

CC APPLICANT: Barry, Gerard F.
CC APPLICANT: Kishore, Ganesh M.
CC APPLICANT: Padgett, Stephen R.
CC APPLICANT: Stallings, William C.
CC TITLE OF INVENTION: Glyphosate Tolerant
CC TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases
CC NUMBER OF SEQUENCES: 69
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F
CC STREET: 700 Chesterfield Village Parkway
CC CITY: St. Louis
CC STATE: Missouri
CC COUNTRY: USA
CC ZIP: 63198
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/833,485
CC FILING DATE: 07-APR-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/306,063
CC FILING DATE: 13-SEP-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/749,611
CC FILING DATE: 28-AUG-1991
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/576,537
CC FILING DATE: 31-AUG-1990
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hoerner Jr., Dennis R.
CC REGISTRATION NUMBER: 30,914
CC REFERENCE/DOCKET NUMBER: 38-21(15117)A
CC TELEPHONE: (314)737-6099
CC TELEFAX: (314)737-6047
CC INFORMATION FOR SEQ ID NO: 67:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 447 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 447 AA; 47046 MW; 944908 CN;
Query Match 86.0%; Score 43; DB 1; Length 447;
Best Local Similarity 71.4%; Pred. No. 3.42e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 393 PLOGAEV 399
QY 347 PLOGGDV 353
RESULT 5
ID US-08-306-063-67 STANDARD; PRT; 447 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 67, Application US/08306063
XX Sequence 67, Application US/08306063
CC Patent No. 5633435
CC GENERAL INFORMATION:
CC APPLICANT: Barry, Gerard F.
CC APPLICANT: Kishore, Ganesh M.

CC APPLICANT: Padgett, Stephen R.
CC APPLICANT: Stallings, William C.
CC TITLE OF INVENTION: Glyphosate Tolerant
CC TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases
CC NUMBER OF SEQUENCES: 69
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F
CC STREET: 700 Chesterfield Village Parkway
CC CITY: St. Louis
CC STATE: Missouri
CC COUNTRY: USA
CC ZIP: 63198
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/306,063
CC FILING DATE: 13-SEP-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/749,611
CC FILING DATE: 28-AUG-1991
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/576,537
CC FILING DATE: 31-AUG-1990
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hoerner Jr., Dennis R.
CC REGISTRATION NUMBER: 30,914
CC REFERENCE/DOCKET NUMBER: 38-21(10660)A
CC TELEPHONE: (314)537-6099
CC TELEFAX: (314)537-6047
CC INFORMATION FOR SEQ ID NO: 67:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 447 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 447 AA; 47046 MW; 944908 CN;
Query Match 86.0%; Score 43; DB 1; Length 447;
Best Local Similarity 71.4%; Pred. No. 3.42e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 393 PLOGAEV 399
QY 347 PLOGGDV 353
RESULT 6
ID US-08-557-309B-37 STANDARD; PRT; 639 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 37, Application US/08557309B
XX Sequence 37, Application US/08557309B
CC Patent No. 5916572
CC GENERAL INFORMATION:
CC APPLICANT: Reed, Steven G.
CC APPLICANT: Skeiky, Yasir A.W.
CC APPLICANT: Lodes, Michael J.
CC APPLICANT: Houghton, Raymond L.
CC TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND
CC NUMBER OF SEQUENCES: 69
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SEED and BERRY LLP

CC STREET: 6300 Columbia Center, 701 Fifth Avenue
CC CITY: Seattle
CC STATE: Washington
CC COUNTRY: USA
CC ZIP: 98104-7092
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC FILING DATE: 14-NOV-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maki, David J.
CC REGISTRATION NUMBER: 31,392
CC REFERENCE/DOCKET NUMBER: 210121.422
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (206) 622-4900
CC TELEFAX: (206) 622-6031
CC INFORMATION FOR SEQ ID NO: 37:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 639 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC SEQUENCE 639 AA; 70748 MW; 1799626 CN;

Query Match 84.0%; Score 42; DB 2; Length 639;
Best Local Similarity 85.7%; Pred No. 4.67e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 415 PLGGDV 421
QY 347 PLGGDV 353

RESULT 7
ID US-09-105-908-8 STANDARD; PRT; 410 AA.

XX xxxxxx

Sequence 8, Application US/09105908

Sequence 8, Application US/09105908
Patent No. 5916793

CC GENERAL INFORMATION:
CC APPLICANT: Filpula, David
CC APPLICANT: Wang, Maoliang
CC TITLE OF INVENTION: Arginine Deiminase Derived From
CC TITLE OF INVENTION: Mycoplasma and Polymer Conjugates Containing the Same
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ROBERTS & MERCANTI
CC STREET: 81 Tamarack Circle
CC CITY: Skillman
CC STATE: New Jersey
CC COUNTRY: United States
CC ZIP: 08558

CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC FILING DATE: 09/09/105.908
CC CLASSIFICATION:
CC PRIOR APPLICATION NUMBER: US/08/792.283

CC FILING DATE: 31-JAN-1997
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Mercanti, Michael N.
CC REGISTRATION NUMBER: 33966
CC REFERENCE/DOCKET NUMBER: 2131055
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 609-921-3500
CC TELEFAX: 609-921-9535
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 410 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: No. 5916793 Relevant
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ORIGINAL SOURCE:
CC ORGANISM: Mycoplasma hominis
CC STRAIN: PG21
CC CELL TYPE: unicellular
CC SEQUENCE 410 AA; 46454 MW; 870644 CN;

Query Match 80.0%; Score 40; DB 2; Length 410;
Best Local Similarity 71.4%; Pred. No. 8.62e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 211 PLGGDV 217
QY 347 PLGGDV 353

RESULT 8

ID US-08-792-283A-8 STANDARD; PRT; 410 AA.

XX xxxxxx

Sequence 8, Application US/08792283A

Sequence 8, Application US/08792283A
Patent No. 5804183

CC GENERAL INFORMATION:
CC APPLICANT: Filpula, David
CC APPLICANT: Wang, Maoliang
CC TITLE OF INVENTION: Arginine Deiminase Derived From
CC TITLE OF INVENTION: Mycoplasma and Polymer Conjugates Containing the Same
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ROBERTS & MERCANTI
CC STREET: 81 Tamarack Circle
CC CITY: Skillman
CC STATE: New Jersey
CC COUNTRY: United States
CC ZIP: 08558
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/792.283A
CC FILING DATE: 31-JAN-1997
CC CLASSIFICATION: 428
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Mercanti, Michael N.
CC REGISTRATION NUMBER: 33966
CC REFERENCE/DOCKET NUMBER: 2131055
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 609-921-3500
CC TELEFAX: 609-921-9535
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 410 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: not relevant
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ORIGINAL SOURCE:
CC ORGANISM: Mycoplasma hominis
CC STRAIN: PG21
CC CELL TYPE: unicellular
CC SEQUENCE 410 AA; 46454 MW; 870644 CN;

Query Match 80.0%; Score 40; DB 1; Length 410;
Best Local Similarity 71.4%; Pred. No. 8.62e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 211 PLOGGDV 217
QY 347 PLOGGDV 353
I:|||||

RESULT 9 STANDARD; PRT; 375 AA.

ID US-08-255-670A-6
XX
AC xxxxxx
XX
DT

Sequence 6, Application US/08255670A

Sequence 6, Application US/08255670A

Patent No. 5691180

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B

APPLICANT: SMITH, PETER L

TITLE OF INVENTION: N-ACETYL-GALACTOSAMINE-TRANSFERASE CDNA

TITLE OF INVENTION: SEQUENCE AND EXPRESSION PRODUCTS

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,

ADDRESSEE: P.C.

STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/255,670A

FILING DATE: 09-JUN-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 5691180man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 2363-092-55

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 375 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 375 AA; 41243 MW; 751891 CN;

Query Match 78.0%; Score 39; DB 1; Length 375;

Best Local Similarity 71.4%; Pred. No. 1.17e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 44 PLOGGEV 50
QY 347 PLOGGDV 353
I:|||||

RESULT 10 STANDARD; PRT; 403 AA.

ID US-08-812-412-2
XX
AC xxxxxx
XX

Sequence 2, Application US/08812412

Sequence 2, Application US/08812412

Patent No. 5989869

GENERAL INFORMATION:

APPLICANT: FILIPPINI, Silvia

APPLICANT: LOMOVSKAYA, Natalia

APPLICANT: FONSTEIN, Leonid

APPLICANT: COLOMBO, Anna L.

APPLICANT: HUTCHINSON, C. R.

TITLE OF INVENTION: PROCESS FOR PREPARING DAUNORUBICIN AND

TITLE OF INVENTION: DOXORUBICIN

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nikaído, Marmelstein, Murray & Oram

STREET: 655 Fifteenth Street N.W. Suite 330

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005-5701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/812,412

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kitts, Monica C.

REGISTRATION NUMBER: 36,105

REFERENCE/DOCKET NUMBER: P1615-7005

TELEPHONE: (202)638-5000

TELEFAX: (202)638-4810

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 403 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 403 AA; 45308 MW; 792488 CN;

Query Match 78.0%; Score 39; DB 2; Length 403;

Best Local Similarity 57.1%; Pred. No. 1.17e+02;

Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 247 PVHGEV 253
QY 347 PLOGGDV 353
I:|||||

RESULT 11 STANDARD; PRT; 251 AA.

ID US-08-152-019A-37

XX

AC xxxxxx

XX

DT

CC TELEX: 910 277299
 CC INFORMATION FOR SEQ ID NO: 2:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 760 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 SQ SEQUENCE 760 AA; 80163 MW; 2975052 CN;
 Query Match 76.0%; Score 38; DB 1; Length 760;
 Best Local Similarity 71.4%; Pred. No. 1.58e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 235 PLDGGAV 241
 Qy 347 PLQGGDV 353

Search completed: Wed May 10 14:23:34 2000
 Job time : 55 secs.

SUMMARY #length 209 #molecular-weight 23114 #checksum 1647

Query Match 88.0%; Score 44; DB 2; Length 209;
 Best Local Similarity 85.7%; Pred. No. 3.58e+00;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 161 PLOGGDL 167
 QY 347 PLOGGDV 353

RESULT 3
 ENTRY #type complete
 TITLE hypF protein - Escherichia coli (strain K-12)
 ORGANISM #formal_name Escherichia coli
 DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change

ACCESSIONS D65051
 REFERENCE A64720
 #authors Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Colliado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

#journal Science (1997) 277:1453-1462
 #title The complete genome sequence of Escherichia coli K-12.
 #cross-references MUID:97426617
 #accession D65051
 #status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA
 #residues 1-750 #label BLAT
 #cross-references GB:AE000355; GB:U00096; NID:92367151; PIDN:AAC75754.1; PID:92367152; UMGp:b2712

#experimental_source strain K-12, substrain MG1655

GENETICS hypF
 #gene #superfamily probabe transcription regulator hypF
 CLASSIFICATION #length 750 #molecular-weight 82065 #checksum 7137
 SUMMARY

Query Match 88.0%; Score 44; DB 2; Length 750;
 Best Local Similarity 85.7%; Pred. No. 3.58e+00;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 490 PLDGDV 496
 QY 347 PLOGGDV 353

RESULT 4
 ENTRY #type complete
 TITLE long chain fatty acyl-CoA synthetase homolog LCFACAS5 - Leishmania major
 ORGANISM #formal_name Leishmania major
 DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change

ACCESSIONS T02837
 REFERENCE Z14740
 #authors Myler, P.J.
 #submission submitted to the EMBL Data Library, May 1998
 #description The nucleotide sequence of Leishmania major Friedlin chromosome 1.

#accession T02837
 #status preliminary; translated from GB/EMBL/DBJ
 #molecule_type DNA
 #residues 1-1607 #label MYL
 #cross-references EMBL:AE001274; NID:93264850; PID:gl617560

GENETICS LCFACAS5
 #gene #map_position 1
 SUMMARY #length 1607 #molecular-weight 168972 #checksum 8040

Query Match 88.0%; Score 44; DB 2; Length 1607;

Best Local Similarity 85.7%; Pred. No. 3.58e+00;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 363 PLOGGSV 369
 QY 347 PLOGGDV 353

RESULT 5
 ENTRY #type complete
 TITLE probabie polyketide synthase - Mycobacterium tuberculosis (strain H37RV)
 ORGANISM #formal_name Mycobacterium tuberculosis
 DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change

ACCESSIONS H70621
 REFERENCE A70500
 #authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, S.; Squares, R.; Sulston, J.E.; Skelton, S.; Squares, S.; Squires, R.; Sultston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

#journal Nature (1998) 393:537-544
 #title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.

#cross-references MUID:98295987
 #accession H70621
 #status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA
 #residues 1-2126 #label COL
 #cross-references GB:285982; GB:AL123456; NID:g3261718; PID:e300651; PID:g1839013

#experimental_source strain H37RV

GENETICS pks7
 #gene #superfamily [acyl-carrier-protein] S-malonyltransferase
 CLASSIFICATION homology; 3-oxoacyl-[acyl-carrier-protein] synthase I
 homology; acyl carrier protein homology; short-chain alcohol dehydrogenase homology

FEATURE 56-456
 563-845
 1694-1875
 1976-2051
 SUMMARY

#domain 3-oxoacyl-[acyl-carrier-protein] synthase I
 homology #label OAS1
 #domain [acyl-carrier-protein] S-malonyltransferase
 homology #label AMT1
 #domain short-chain alcohol dehydrogenase homology
 #label SADH
 #domain acyl carrier protein homology #label ACP2
 #length 2126 #molecular-weight 221045 #checksum 2737

Query Match 88.0%; Score 44; DB 2; Length 2126;
 Best Local Similarity 71.4%; Pred. No. 3.58e+00;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1222 PLOGGDI 1228
 QY 347 PLOGGDV 353

RESULT 6
 ENTRY #type complete
 TITLE hypF1 protein - Alcaligenes eutrophus
 ORGANISM #formal_name Alcaligenes eutrophus
 DATE 06-Jan-1995 #sequence_revision 24-Oct-1997 #text_change

ACCESSIONS S29976
 REFERENCE S62600
 #authors Darnedde, J.; Eitinger, T.; Patenge, N.; Friedrich, B.

```
#journal
#title
#cross-references EMBL:235351-358
#accession S62601
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-394 #label DER
#cross-references EMBL:X70183; NID:g577727; PID:g577728
#experimental_source strain H16
#note this is a revision to the sequence from reference S29975
REFERENCE
S29975
#authors Eitinger, T.
#submission submitted to the EMBL Data Library, January 1993
#accession S29976
#status preliminary
#molecule_type DNA
#residues
'/MCVICGNTNHTARODENKGEHAPGRLVDTGTPPGAHHVRIDASTGDLHY
GAGPAHVPGLSOARAIKLEQDVGLGNHNNRAHNRQHFVAGVLAALNVSSPG
SGKTLTCTTEALRRCADLQIAVTEGDOOTSQSHDAERIRATGVPAIQINIGK
CHLDALMVANAYERLPLHAAHAHTEHREOEDGSHSHHHDHHRDHDHRS
GIGSVLFIEVNGVLCVPMWDLGESAKVALLSVTEGEDKPLKYPDFEAAASLM
LNKIDLLPHLRFVAVRCIEYARQVNPQLVLSAATGEGVDEWDLWLAGGAA
AGAAAATDGRMAIAAPEPSLPSERSWLPVRRRCRCE',1-300,'I',
302-394 #label EIT
#cross-references EMBL:X70183
#note this sequence has been revised in reference S62600
REFERENCE
I39528
#authors Deredde, J.; Eitinger, M.; Friedrich, B.
#journal Arch. Microbiol. (1993) 159:545-553
#title Analysis of a pleiotropic gene region involved in formation
of catalytically active hydrogenases in Alcaligenes
eutrophus H16
#cross-references MUID:93356597
#accession I39530
#status preliminary; translated from GE/EMBL/DDBJ
#molecule_type DNA
#residues 1-300,'I',302-394 #label RES
#cross-references EMBL:X70183; NID:g577727; PID:g577728
GENETICS
#gene hypF
#keywords DNA binding; transcription regulation
SUMMARY
#length 394 #molecular-weight 40677 #checksum 1632
Query Match 86.0%; Score 43; DB 2; Length 394;
Best Local Similarity 85.7%; Pred. No. 6.31e+00;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 157 PLPGGDV 163
|||
QY 347 PLOGGDV 353
|||
RESULT 7
ENTRY #type complete
TITLE portal protein gp3 - phage HK97
ORGANISM #formal_name phage HK97
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
10-Sep-1997
ACCESSIONS S54390
REFERENCE S54390
#authors Duda, R.L.; Martincic, K.; Hendrix, R.W.
#journal J. Mol. Biol. (1995) 247:636-647
#title Genetic basis of bacteriophage HK97 prohead assembly.
#cross-references MUID:95239740
#accession S54390
#molecule_type DNA
#residues 1-424 #label DUD
#cross-references EMBL:U18319; NID:g609309; PID:g609310
GENETICS
#gene
#keywords ABC transporter
SUMMARY
#length 424 #molecular-weight 47266 #checksum 2285
```

```
Query Match 86.0%; Score 43; DB 2; Length 424;
Best Local Similarity 85.7%; Pred. No. 6.31e+00;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 394 PLPGGDV 400
|||
QY 347 PLOGGDV 353
|||
RESULT 8
ENTRY #type complete
TITLE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) -
Synecocystis sp. (strain PCC 6803)
ALTERNATE_NAMES 5-enolpyruvylshikimate 3-phosphate synthase
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
18-Jun-1999
ACCESSIONS S76218; S33748; S34393
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synecocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S76218
#molecule_type DNA
#residues 1-447 #label KAN
#cross-references EMBL:D90914; GB:AB001339; NID:gi653477;
PIDN:BAAL8477.1; PID:dl019210; PID:gl653564
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
REFERENCE S33748
#authors Mayes, S.R.; dalla Chiesa, M.; Zhang, Z.; Barber, J.
#journal FEBS Lett. (1993) 325:255-261
#title The genes araA and trnQ are located upstream of psbO in the
chromosome of Synecocystis 6803.
#cross-references MUID:93307506
#accession S33748
#molecule_type DNA
#residues 103-194 #label MAY
#cross-references EMBL:X72784; NID:g394747; PIDN:CAA51291.1;
PID:g394748
GENETICS
#gene araA
CLASSIFICATION #superfamily 3-phosphoshikimate 1-carboxyvinyltransferase;
3-phosphoshikimate 1-carboxyvinyltransferase homology
aromatic amino acid biosynthesis; transferase
KEYWORDS
FEATURE
28-439 #domain 3-phosphoshikimate 1-carboxyvinyltransferase
homology #label PSK
SUMMARY #length 447 #molecular-weight 47046 #checksum 2349
Query Match 86.0%; Score 43; DB 2; Length 447;
Best Local Similarity 71.4%; Pred. No. 6.31e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 393 PLOGAEV 399
|||
QY 347 PLOGGDV 353
|||
RESULT 9
ENTRY #type complete
TITLE ABC transporter sl10778 - Synecocystis sp. (strain PCC 6803)
```

```

ALTERNATE_NAMES protein sl10778
ORGANISM #formal_name Synecocystis sp.
#variety
DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
S77032
S74322
ACCESSIONS S77032
REFERENCE S74322
#authors
Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Nariu, K.; Okumura, S.; Shimo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synecocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S77032
#molecule_type DNA
#status nucleic acid sequence not shown; translation not shown
#residues 1-790 #label KAN
#cross-references EMBL:D64005; GB:AB001339; NID:gl001779; PID:dl011375;
PID:gl006575
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
CLASSIFICATION #superfamily Synecocystis ABC transporter sl10778;
ATP-binding cassette homology
KEYWORDS ATP; P-loop; transport protein
FEATURE
244-434 #domain ATP-binding cassette homology #label ABC\
261-468 #region nucleotide-binding motif A (P-loop)
SUMMARY #length 790 #molecular-weight 87656 #checksum 7314
Query Match 86.0%; Score 43; DB 1; Length 790;
Best Local Similarity 85.7%; Pred. No. 6.31e+00;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 93 PLOGGDV 99
Qy 347 PLOGGDV 353
RESULT 10
ENTRY #type complete
TITLE Xin protein - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
ACCESSIONS T14266
REFERENCE T14266
#authors Wang, D.-Z.; Lin, J.J.C.
#submission submitted to the EMBL Data Library, March 1998
#description Involvement of a novel gene, Xin, in cardiac looping.
#accession T14266
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-2562 #label WAN
#cross-references EMBL:AF051944; NID:g2970643; PID:g2970644;
#experimental_source cardiac muscle
SUMMARY #length 2562 #molecular-weight 284482 #checksum 2286
Query Match 86.0%; Score 43; DB 2; Length 2562;
Best Local Similarity 85.7%; Pred. No. 6.31e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 211 PVOGGDV 217
Qy 347 PLOGGDV 353

```

```

RESULT 11
ENTRY #type complete
TITLE hypothetical protein 023L - Chilo iridescent virus
ORGANISM #formal_name Chilo iridescent virus
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
07-May-1999
ACCESSIONS T03052
REFERENCE Z14834
#authors Bahr, U.; Tidona, C.A.; Darai, G.
#journal Virus Genes (1997) 15:235-245
#title The DNA sequence of Chilo iridescent virus between the genome
coordinates 0.101 and 0.391; similarities in coding
strategy between insect and vertebrate iridoviruses.
#cross-references MUID:98141693
#accession T03052
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-165 #label BAH
#cross-references EMBL:AF003534; NID:g2738385; PID:g2738399
SUMMARY #length 165 #molecular-weight 18412 #checksum 5977
Query Match 84.0%; Score 42; DB 2; Length 165;
Best Local Similarity 71.4%; Pred. No. 1.10e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 124 PLPGGEV 130
Qy 347 PLOGGDV 353
RESULT 12
ENTRY #type complete
TITLE hypothetical protein Ape1597 - Aeropyrum pernix (strain K1)
ORGANISM #formal_name Aeropyrum pernix
DATE 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
ACCESSIONS H72538
REFERENCE A72450
#authors Kavarabavasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.;
Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba,
S.; Ankei, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai,
Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.;
Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura,
Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
#journal DNA Res. (1999) 6:83-101
#title Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
#cross-references MUID:99310639
#accession H72538
#status preliminary
#molecule_type DNA
#residues 1-175 #label KAW
#cross-references DBJ:AP000062; NID:g5105244; PIDN:BAA80597.1;
PID:dl044383; PID:g5105284
#experimental_source strain K1
GENETICS
#gene APE1597
#summary #length 175 #molecular-weight 18953 #checksum 9362
Query Match 84.0%; Score 42; DB 2; Length 175;
Best Local Similarity 83.3%; Pred. No. 1.10e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 158 PLOGGE 163
Qy 347 PLOGGD 352
RESULT 13
ENTRY #type complete
TITLE probable cell division protein - syphilis spirochete
ORGANISM #formal_name Treponema pallidum subsp. pallidum #common_name

```

```

syphilis spirochete
24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
17-Mar-1999
ACCESSIONS H71255
REFERENCE A71250
#authors Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Winn, M.; Hickey, E.K.; Clayton, R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod, M.P.; Salzberg, S.; Peterson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald, L.; Artiach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.; Garland, S.; Horst, K.; Roberts, K.; Matthey, J.C.; Weidman, J.; Smith, H.O.; Venter, J.C.
#journal Science (1998) 281:375-388
#title Complete genome sequence of Treponema pallidum, the syphilis spirochete.
#cross-references MUID:98332770
#accession H71255
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-799 #label COL
##cross-references GB:AE001267; GB:AE000520; NID:g3323319; PID:g3323325
##experimental_source strain Nichols
GENETICS
#gene TP0999
SUMMARY
#length 799 #molecular-weight 86605 #checksum 4462
Query Match 84.0%; Score 42; DB 2; Length 799;
Best Local Similarity 83.3%; Pred. No. 1.10e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 212 PLOGGE 217
QY 347 PLOGGD 352
|||||

RESULT 14
ENTRY S64929 #type complete
TITLE hypothetical protein YLR095c - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein L2542
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 01-Aug-1995 #sequence_revision 24-May-1996 #text_change
14-Nov-1997
ACCESSIONS S64929
REFERENCE S64920
#authors Benes, V.; Rechmann, S.; Nentwich, U.; Schwager, C.; Ansoorge, W.; Voss, H.
#submission submitted to the Protein Sequence Database, May 1996
#accession S64929
##molecule_type DNA
##residues 1-812 #label BEN
##cross-references EMBL:273267; NID:gr1360472; PID:e245800; PID:gr1360473; MIPS:YLR095c
##experimental_source strain S288C
GENETICS
#map_position 12R
SUMMARY
#length 812 #molecular-weight 93445 #checksum 1139
Query Match 84.0%; Score 42; DB 2; Length 812;
Best Local Similarity 85.7%; Pred. No. 1.10e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 501 PLOGGDV 507
QY 347 PLOGGDV 353
|||||

RESULT 15
ENTRY B70199 #type complete
TITLE outer membrane protein homolog - Lyme disease spirochete
ORGANISM #formal_name Borrelia burgdorferi #common_name Lyme disease

```

```

spirochete
13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
05-Jun-1998
ACCESSIONS B70199
REFERENCE A70100
#authors Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Winn, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Utterback, T.; Matthey, L.; McDonald, L.; Artiach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
#journal Nature (1997) 390:580-586
#title Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
#cross-references MUID:98065943
#accession B70199
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-821 #label KLE
##cross-references GB:AE001178; GB:AE000783; NID:g2688731; PID:g2688732; TIGR:BB0795
##experimental_source strain B31
SUMMARY
#length 821 #molecular-weight 94567 #checksum 3345
Query Match 84.0%; Score 42; DB 2; Length 821;
Best Local Similarity 85.7%; Pred. No. 1.10e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 379 PLOGGDV 385
QY 347 PLOGGDV 353
|||||

Search completed: Wed May 10 14:21:57 2000
Job time : 8 secs.

```

W P S R E L I
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 13:57:41 2000; MasPar time 47.15 seconds
Tabular output not generated. 2.198 Million cell updates/sec

Title: >US-09-376-430-2
Description: (286-293) from US09376430A.pep (21 of 25)
Perfect Score: 60
Sequence: 1 WITDTQNV 8

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 14.978; Variance 57.012; scale 0.263

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	45	75.0	265	2	US-07-857- Sequence 66, Applicati	3.10e+02
2	45	75.0	379	1	US-08-227- Sequence 18, Applicati	3.10e+02
3	45	75.0	379	1	US-09-073- Sequence 18, Applicati	3.10e+02
4	45	75.0	580	2	US-08-468- Sequence 69, Applicati	3.10e+02
5	45	75.0	580	1	US-08-414- Sequence 69, Applicati	3.10e+02
6	45	75.0	580	2	US-08-486- Sequence 69, Applicati	3.10e+02
7	45	75.0	1255	1	US-08-414- Sequence 68, Applicati	3.10e+02
8	45	75.0	1255	2	US-08-484- Sequence 8, Applicati	3.10e+02
9	45	75.0	1255	2	US-08-356- Sequence 2, Applicati	3.10e+02
10	45	75.0	1255	1	US-08-467- Sequence 68, Applicati	3.10e+02
11	45	75.0	1255	2	US-08-468- Sequence 68, Applicati	3.10e+02
12	45	75.0	1255	2	US-08-625- Sequence 2, Applicati	3.10e+02
13	45	75.0	1255	2	US-08-486- Sequence 68, Applicati	3.10e+02
14	44	73.3	1968	2	US-08-223- Sequence 45, Applicati	3.80e+02
15	44	73.3	1968	1	US-08-455- Sequence 45, Applicati	3.80e+02
16	44	73.3	1968	2	US-08-311- Sequence 7, Applicati	3.80e+02
17	44	73.3	1968	1	US-07-745- Sequence 7, Applicati	3.80e+02
18	44	73.3	2161	1	US-07-745- Sequence 2, Applicati	3.80e+02
19	44	73.3	2161	1	US-08-455- Sequence 49, Applicati	3.80e+02
20	44	73.3	2161	1	US-08-455- Sequence 51, Applicati	3.80e+02
21	44	73.3	2161	2	US-08-223- Sequence 49, Applicati	3.80e+02
22	44	73.3	2161	2	US-08-311- Sequence 2, Applicati	3.80e+02
23	44	73.3	2161	2	US-08-223- Sequence 51, Applicati	3.80e+02

24	44	73.3	2366	2	US-08-405- Sequence 10, Applicati	3.80e+02
25	44	73.3	2366	1	US-08-480- Sequence 10, Applicati	3.80e+02
26	42	70.0	15	3	PCT-US93-0 Sequence 16, Applicati	5.68e+02
27	42	70.0	106	2	US-08-850- Sequence 12, Applicati	5.68e+02
28	42	70.0	106	1	US-08-241- Sequence 12, Applicati	5.68e+02
29	42	70.0	166	3	PCT-US93-0 Sequence 9, Applicatio	5.68e+02
30	42	70.0	166	3	PCT-US93-0 Sequence 4, Applicatio	5.68e+02
31	42	70.0	166	3	PCT-US93-0 Sequence 8, Applicatio	5.68e+02
32	42	70.0	166	3	PCT-US93-0 Sequence 10, Applicati	5.68e+02
33	42	70.0	264	2	US-07-857- Sequence 67, Applicati	5.68e+02
34	42	70.0	280	2	US-08-826- Sequence 12, Applicati	5.68e+02
35	42	70.0	280	2	US-08-820- Sequence 12, Applicati	5.68e+02
36	42	70.0	280	2	US-08-424- Sequence 12, Applicati	5.68e+02
37	42	70.0	288	2	US-08-820- Sequence 11, Applicati	5.68e+02
38	42	70.0	288	2	US-08-826- Sequence 11, Applicati	5.68e+02
39	42	70.0	288	2	US-08-424- Sequence 11, Applicati	5.68e+02
40	42	70.0	333	2	US-08-853- Sequence 48, Applicati	5.68e+02
41	41	68.3	428	1	US-07-882- Sequence 4, Applicatio	6.93e+02
42	41	68.3	570	2	US-08-484- Sequence 16, Applicati	6.93e+02
43	41	68.3	570	2	US-08-480- Sequence 16, Applicati	6.93e+02
44	41	68.3	570	2	US-08-484- Sequence 16, Applicati	6.93e+02
45	41	68.3	570	2	US-08-484- Sequence 16, Applicati	6.93e+02

ALIGNMENTS

RESULT 1
ID US-07-857-224B-66 STANDARD; PRT; 265 AA.
XX
AC
XX
XX
DT
XX
DE
XX
Sequence 66, Application US/07857224B
Sequence 66, Application US/07857224B
Patent No. 5958784
GENERAL INFORMATION:
APPLICANT: Benner, Steven A.
TITLE OF INVENTION: Predicting Folded Structures of Proteins
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven A. Benner
STREET: Hadlaubstrasse 151
CITY: Zurich
STATE: none
COUNTRY: Switzerland
ZIP: (note: this is an international post code) CH-8092
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/857,224B
FILING DATE: 03/25/92
CLASSIFICATION: 436
PRIOR APPLICATION DATA: none
TELECOMMUNICATION INFORMATION:
TELEPHONE: (International) 41 1 632 2830
TELEFAX: (International) 41 1 262 2437
TELEX: none
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 265
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: protein
ORIGINAL SOURCE: human
ORGANISM: human
FEATURE: Protein kinase; Table 8 Column 75
PUBLICATION INFORMATION:

CC AUTHORS: Hanks, S. K.
CC AUTHORS: Quinn, A. M.
CC AUTHORS: Hunter, T.
CC TITLE: The protein kinase family
CC JOURNAL: Science
CC VOLUME: 241
CC PAGES: 42-52
CC DATE: 1988
CC SEQUENCE 265 AA; 30122 MW; 367139 CN;

Query Match 75.0%; Score 45; DB 2; Length 265;
Best Local Similarity 62.5%; Pred. No. 3.10e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 22 WIPDGENV 29
QY 286 WITDTQNV 293

RESULT 2
ID US-08-227-108-18 STANDARD; PRT; 379 AA.
XX
AC xxxxxx
XX
DT
DE
DE Sequence 18, Application US/08227108
XX

Sequence 18, Application US/08227108
Patent No. 5807726
GENERAL INFORMATION:
APPLICANT: Blanchard, Claire
APPLICANT: Benicourt, Claude
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: Recombinant Dog Gastric Lipase
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/227,108
FILING DATE: 03-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fanucci, Allan A.
REGISTRATION NUMBER: 30,256
REFERENCE/DOCKET NUMBER: 7620-033
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 379 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 379 AA; 42909 MW; 769626 CN;

Query Match 75.0%; Score 45; DB 1; Length 379;
Best Local Similarity 50.0%; Pred. No. 3.10e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 22 WIPDGENV 29
QY 286 WITDTQNV 293

Db 325 WLADPDV 332
QY 286 WITDTQNV 293

RESULT 3
ID US-09-073-674-18 STANDARD; PRT; 379 AA.
XX
AC xxxxxx
XX
DT
DE
DE Sequence 18, Application US/09073674
XX

Sequence 18, Application US/09073674
Patent No. 5998189
GENERAL INFORMATION:
APPLICANT: Blanchard, Claire
APPLICANT: Benicourt, Claude
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: Recombinant Dog Gastric Lipase
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Warner-Lambert Company
STREET: 2800 Plymouth Road
CITY: Ann Arbor
STATE: Michigan
COUNTRY: U.S.A.
ZIP: 48105

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073,674
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Crissey, Todd M.
REGISTRATION NUMBER: 37,807
REFERENCE/DOCKET NUMBER: 5072-D1-66-TMC
TELEPHONE: 734 622-7530
TELEFAX: 734 622-1553
TELEX:
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 379 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 379 AA; 42909 MW; 769626 CN;

Query Match 75.0%; Score 45; DB 2; Length 379;
Best Local Similarity 50.0%; Pred. No. 3.10e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 325 WLADPDV 332
QY 286 WITDTQNV 293

RESULT 4
ID US-08-468-545B-69 STANDARD; PRT; 580 AA.
XX
AC xxxxxx
XX
DT
DE
DE Sequence 69, Application US/08468545B

XX Sequence 69, Application US/08468545B
CC Patent No. 5876712
CC GENERAL INFORMATION:
CC APPLICANT: Cheever, Martin A.
CC APPLICANT: Disis, Mary L.
CC TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
CC TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
CC TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
CC NUMBER OF SEQUENCES: 69
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Seed and Berry LLP
CC STREET: 6300 Columbia Center, 701 Fifth Avenue
CC CITY: Seattle
CC STATE: Washington
CC COUNTRY: US
CC ZIP: 98104-7092
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/468,545B
CC FILING DATE: 06-JUN-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Sharkey, Richard G.
CC REGISTRATION NUMBER: 32,629
CC REFERENCE/DOCKET NUMBER: 920010.448C5
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (206) 622-4900
CC TELEFAX: (206) 682-6031
CC INFORMATION FOR SEQ ID NO: 69:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 580 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC SEQUENCE 580 AA; 64054 MW; 1767456 CN;
Query Match 75.0%; Score 45; DB 2; Length 580;
Best Local Similarity 62.5%; Pred. No. 3.10e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 64 WIPDGENV 71
QY ||| : ||
286 WITDTQNV 293
RESULT 5
ID US-08-414-417B-69 STANDARD; PRT; 580 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 69, Application US/08414417B
XX Sequence 69, Application US/08414417B
CC Patent No. 5801005
CC GENERAL INFORMATION:
CC APPLICANT: Cheever, Martin A.
CC APPLICANT: Disis, Mary L.
CC TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
CC TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
CC TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
CC NUMBER OF SEQUENCES: 69
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Seed and Berry LLP
CC STREET: 6300 Columbia Center, 701 Fifth Avenue
CC CITY: Seattle
CC STATE: Washington

CC COUNTRY: US
CC ZIP: 98104-7092
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/414,417B
CC FILING DATE: 31-MAR-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Sharkey, Richard G.
CC REGISTRATION NUMBER: 32,629
CC REFERENCE/DOCKET NUMBER: 920010.448C2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (206) 622-4900
CC TELEFAX: (206) 682-6031
CC INFORMATION FOR SEQ ID NO: 69:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 580 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC SEQUENCE 580 AA; 64054 MW; 1767456 CN;
Query Match 75.0%; Score 45; DB 1; Length 580;
Best Local Similarity 62.5%; Pred. No. 3.10e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 64 WIPDGENV 71
QY ||| : ||
286 WITDTQNV 293
RESULT 6
ID US-08-486-348A-69 STANDARD; PRT; 580 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 69, Application US/08486348A
XX Sequence 69, Application US/08486348A
CC Patent No. 5846538
CC GENERAL INFORMATION:
CC APPLICANT: Cheever, Martin A.
CC APPLICANT: Disis, Mary L.
CC TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
CC TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH TH
CC TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
CC NUMBER OF SEQUENCES: 69
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Seed and Berry LLP
CC STREET: 6300 Columbia Center, 701 Fifth Avenue
CC CITY: Seattle
CC STATE: Washington
CC COUNTRY: US
CC ZIP: 98104-7092
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/486,348A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Sharkey, Richard G.
CC REGISTRATION NUMBER: 32,629
CC REFERENCE/DOCKET NUMBER: 920010.448C6

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (206) 622-4900

CC TELEFAX: (206) 682-6031

CC INFORMATION FOR SEQ ID NO: 69:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 580 amino acids

CC TYPE: amino acid

CC STRANDEDNESS:

CC TOPOLOGY: linear

CC SEQUENCE 580 AA; 64054 MW; 1767456 CN;

Query Match

Best Local Similarity 75.0%; Score 45; DB 2; Length 580;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 64 WIPDGENV 71

||| :||

QY 286 WITDQNV 293

RESULT 7

ID US-08-414-417B-68

STANDARD; PRT; 1255 AA.

XX xxxxxx

AC

XX

DT

DE

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (206) 622-4900

CC TELEFAX: (206) 682-6031

CC INFORMATION FOR SEQ ID NO: 69:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 580 amino acids

CC TYPE: amino acid

CC STRANDEDNESS:

CC TOPOLOGY: linear

CC SEQUENCE 580 AA; 64054 MW; 1767456 CN;

Query Match

Best Local Similarity 75.0%; Score 45; DB 2; Length 580;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 64 WIPDGENV 71

||| :||

QY 286 WITDQNV 293

RESULT 7

ID US-08-414-417B-68

STANDARD; PRT; 1255 AA.

XX xxxxxx

AC

XX

DT

DE

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (206) 622-4900

CC TELEFAX: (206) 682-6031

CC INFORMATION FOR SEQ ID NO: 69:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 580 amino acids

CC TYPE: amino acid

CC STRANDEDNESS:

CC TOPOLOGY: linear

CC SEQUENCE 580 AA; 64054 MW; 1767456 CN;

Query Match

Best Local Similarity 75.0%; Score 45; DB 2; Length 580;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 64 WIPDGENV 71

||| :||

QY 286 WITDQNV 293

RESULT 7

ID US-08-414-417B-68

STANDARD; PRT; 1255 AA.

XX xxxxxx

AC

XX

DT

DE

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (206) 622-4900

CC TELEFAX: (206) 682-6031

CC INFORMATION FOR SEQ ID NO: 69:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 580 amino acids

CC TYPE: amino acid

CC STRANDEDNESS:

CC TOPOLOGY: linear

CC SEQUENCE 580 AA; 64054 MW; 1767456 CN;

Query Match

Best Local Similarity 75.0%; Score 45; DB 2; Length 580;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 64 WIPDGENV 71

||| :||

QY 286 WITDQNV 293

RESULT 7

ID US-08-414-417B-68

STANDARD; PRT; 1255 AA.

XX xxxxxx

AC

XX

DT

DE

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (206) 622-4900

CC TELEFAX: (206) 682-6031

CC INFORMATION FOR SEQ ID NO: 69:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 580 amino acids

CC TYPE: amino acid

CC STRANDEDNESS:

CC TOPOLOGY: linear

CC SEQUENCE 580 AA; 64054 MW; 1767456 CN;

Query Match

Best Local Similarity 75.0%; Score 45; DB 2; Length 580;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 64 WIPDGENV 71

||| :||

QY 286 WITDQNV 293

RESULT 7

ID US-08-414-417B-68

STANDARD; PRT; 1255 AA.

XX xxxxxx

AC

XX

DT

DE

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC TELECOMMUNICATION INFORMATION:

QY 286 WITDTQNV 293
|| | : ||

RESULT 9
ID US-08-356-786-2 STANDARD: PRT: 1255 AA.

XX
AC xxxxxx

DE Sequence 2, Application US/08356786

XX Sequence 2, Application US/08356786

CC Patent No. 5877305

CC GENERAL INFORMATION:

CC APPLICANT: Huston, James S.

CC APPLICANT: Oppermann, Hermann

CC APPLICANT: Houston, L. L.

CC APPLICANT: Ring, David B.

CC TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer

CC TITLE OF INVENTION: Marker

CC NUMBER OF SEQUENCES: 16

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault

CC STREET: Exchange Place, 53 State Street

CC CITY: Boston

CC STATE: Massachusetts

CC COUNTRY: USA

CC ZIP: 02109

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC SOFTWARE: PatentIn Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/356,786

CC FILING DATE:

CC CLASSIFICATION: 424

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 07/831,967

CC FILING DATE: 06-FEB-1992

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Pitcher, Edmund R.

CC REGISTRATION NUMBER: 27,829

CC REFERENCE/DOCKET NUMBER: CRP-053

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 248-7000

CC TELEFAX: (617) 248-7100

CC INFORMATION FOR SEQ ID NO: 2:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 1255 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 1255 AA; 137909 MW; 8111405 CN;

Query Match 75.0%; Score 45; DB 2; Length 1255;
Best Local Similarity 62.5%; Pred. No. 3.10e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 739 WIPDGENV 746

|| | : ||

QY 286 WITDTQNV 293

RESULT 10

ID US-08-467-083-68 STANDARD: PRT: 1255 AA.

XX

AC xxxxxx

XX

DT

XX

DE Sequence 68, Application US/08467083

XX

CC Sequence 68, Application US/08467083

CC Patent No. 5726023

CC GENERAL INFORMATION:

CC APPLICANT: Cheever, Martin A.

CC APPLICANT: Disis, Mary L.

CC TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/NEU PROTEIN

CC TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH TH

CC TITLE OF INVENTION: HER-2/NEU ONCOGENE IS ASSOCIATED

CC NUMBER OF SEQUENCES: 68

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Seed and Berry

CC STREET: 6300 Columbia Center, 701 Fifth Avenue

CC CITY: Seattle

CC STATE: Washington

CC COUNTRY: US

CC ZIP: 98104-7092

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: PatentIn Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/467,083

CC FILING DATE: 06-JUN-1995

CC CLASSIFICATION: 424

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/414,417

CC FILING DATE: 06-JUN-1995

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Sharkey, Richard G.

CC REGISTRATION NUMBER: 32,629

CC REFERENCE/DOCKET NUMBER: 920010.448C2

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (206) 622-4900

CC TELEFAX: (206) 682-6031

CC TELEX: 3723836 SEEDANBERRY

CC INFORMATION FOR SEQ ID NO: 68:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 1255 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC SEQUENCE 1255 AA; 137955 MW; 8109851 CN;

Query Match 75.0%; Score 45; DB 1; Length 1255;
Best Local Similarity 62.5%; Pred. No. 3.10e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 739 WIPDGENV 746

|| | : ||

QY 286 WITDTQNV 293

RESULT 11

ID US-08-468-545B-68 STANDARD: PRT: 1255 AA.

XX

AC xxxxxx

XX

DT

XX

DE

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Sequence 68, Application US/08468545B

Sequence 68, Application US/08468545B

Patent No. 5876712

GENERAL INFORMATION:

APPLICANT: Cheever, Martin A.

APPLICANT: Disis, Mary L.

TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/NEU PROTEIN

TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH TH

TITLE OF INVENTION: HER-2/NEU ONCOGENE IS ASSOCIATED

NUMBER OF SEQUENCES: 69

CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Seed and Berry LLP
CC STREET: 6300 Columbia Center, 701 Fifth Avenue
CC CITY: Seattle
CC STATE: Washington
CC COUNTRY: US
CC ZIP: 98104-7092
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/468,545B
CC FILING DATE: 06-JUN-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Sharkey, Richard G.
CC REGISTRATION NUMBER: 32,629
CC REFERENCE/DOCKET NUMBER: 920010.448C5
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (206) 622-4900
CC TELEFAX: (206) 682-6031
CC INFORMATION FOR SEQ ID NO: 68:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1255 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 1255 AA; 137955 MW; 8109851 CN;

Query Match 75.0%; Score 45; DB 2; Length 1255;
Best Local Similarity 62.5%; Pred. No. 3.10e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 739 WIPDGENV 746
||| :||
QY 286 WITDTQNV 293

RESULT 12

ID US-08-625-101-2 STANDARD; PRT; 1255 AA.

AC xxxxxx

Sequence 2, Application US/08625101

Sequence 2, Application US/08625101

Patent No. 5859445

GENERAL INFORMATION:

APPLICANT: Cheever, Martin A.

APPLICANT: Disis, Mary L.

TITLE OF INVENTION: COMPOUNDS FOR ELICITING OR ENHANCING IMMUNE

TITLE OF INVENTION: REACTIVITY TO HER-2/neu PROTEIN FOR PREVENTION

TITLE OF INVENTION: OR TREATMENT OF MALIGNANCIES IN WHICH THE HER-2/neu

TITLE OF INVENTION: ONCOGENE IS ASSOCIATED

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/625,101

FILING DATE: 01-APR-1996

CLASSIFICATION: 424

CC ATTORNEY/AGENT INFORMATION:
CC NAME: Sharkey, Richard G.
CC REGISTRATION NUMBER: 32,629
CC REFERENCE/DOCKET NUMBER: 920010.448C7
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (206) 622-4900
CC TELEFAX: (206) 682-6031
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1255 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 1255 AA; 137909 MW; 8111405 CN;

Query Match 75.0%; Score 45; DB 2; Length 1255;
Best Local Similarity 62.5%; Pred. No. 3.10e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 739 WIPDGENV 746

||| :||

QY 286 WITDTQNV 293

RESULT 13

ID US-08-486-348A-68 STANDARD; PRT; 1255 AA.

XX xxxxxx

AC xxxxxx

DT xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

XX xxxxxx

```
Best Local Similarity 62.5%; pred. No. 3.10e+02;
5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

DB 739 WIPDGENV 746
|||: ||
OV 286 WITDTQNV 293

RESULT	14	
ID	US-08-223-305C-45	STANDARD;
XX		PRT; 1968 AA.
AC	xxxxxx	

Sequence 45, Application US/08223305C

Sequence 45, Application US/08223305C
Patent No. 5851824

Patent No. 2,601,024

GENERAL INFORMATION:

APPLICANT: Harpold, Michael

APPLICANT: Ellis, Steven

APPLICANT: Williams, Mark

APPLICANT: Feldman, Daniel

APPLICANT: McCue, Ann

APPLICANT: Brenner, Robert

TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND METHODS

NUMBER OF SEQUENCES: 57

CC
CC
CC
CC
CC
CC
CC
CC
CC
CC

NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Hallier & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: California

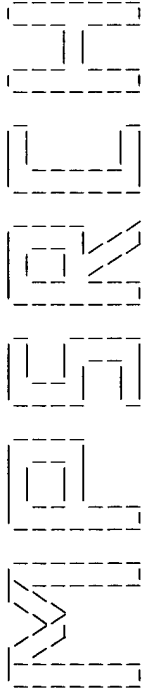
CC SERIAL: CARLLOLUNA
CC COUNTRY: USA
CC ZIP: 92101-2926
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSEQ version 1.5
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/223.305C
CC FILING DATE: April 4, 1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/868.354
CC FILING DATE: April 10, 1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/745.206
CC FILING DATE: 15-AUG-1991
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/620.250
CC FILING DATE: 30-NOV-1990
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/482.384
CC FILING DATE: 20-FEB-1990
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/603.751
CC FILING DATE: 04-APR-1989
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: WO PCT/US89/01408
CC FILING DATE: 04-APR-1989
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/176.899
CC FILING DATE: 04-APR-1988
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Seidman, Stephanie L.
CC REGISTRATION NUMBER: 33,779
CC REFERENCE/DOCKET NUMBER: 52516 (P519)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619)238-0999
CC TELEFAX: (619)238-0062
CC INFORMATION FOR SEQ ID NO: 45:

CC APPLICATION NUMBER: WO PCT/US89/01408
CC FILING DATE: 04-APR-1989
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/176,899
CC FILING DATE: 04-APR-1988
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Seidman, Stephanie L.
CC REGISTRATION NUMBER: 33,779
CC REFERENCE/DOCKET NUMBER: 6362-52517
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619)238-0999
CC TELEFAX: (619)238-0062
CC INFORMATION FOR SEQ ID NO: 45:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1968 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC FRAGMENT TYPE: internal
CC SQ SEQUENCE 1968 AA; 22287 MW; 19913781 CN;

Query Match 73.3%; Score 44; DB 1; Length 1968;
Best Local Similarity 37.5%; Pred. No. 3.80e+02;
Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 440 WITQAEI 447
QY 286 WITDTQNV 293
|||||

Search completed: Wed May 10 13:58:37 2000
Job time : 56 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 13:56:51 2000; MasPar time 3.93 Seconds
Tabular output not generated. 96.085 Million cell updates/sec

Title: >US-09-376-430-2
Description: (286-293) from US09376430A.pep (21 of 25)
Perfect Score: 60
Sequence: 1 WITDTQNV 8

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 22.537; Variance 30.764; scale 0.733

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	48	80.0	342	2 H70653	hypothetical protein	2.35e+01
2	48	80.0	626	2 T09345	hypothetical protein	2.35e+01
3	46	76.7	286	2 D75217	probable 2-acetyl-1-a	5.42e+01
4	46	76.7	533	2 S33701	homeotic protein Dth-	5.42e+01
5	46	76.7	583	1 A41129	radixin - mouse	5.42e+01
6	46	76.7	1014	2 S75724	hypothetical protein	5.42e+01
7	46	76.7	2204	1 RRNZNV	genome polyprotein tra	5.42e+01
8	45	75.0	281	2 A72472	probable electron tra	8.16e+01
9	45	75.0	455	2 T16070	hypothetical protein	8.16e+01
10	45	75.0	481	2 F70890	hypothetical protein	8.16e+01
11	45	75.0	661	2 S19646	DNA repair protein SN	8.16e+01
12	45	75.0	974	2 S15038	cell division control	8.16e+01
13	45	75.0	1075	2 S54067	probable membrane pro	8.16e+01
14	45	75.0	1254	2 I48161	p-185 precursor - gol	8.16e+01
15	45	75.0	1255	1 A24571	protein-tyrosine kina	8.16e+01
16	45	75.0	1260	1 TVRTNU	protein-tyrosine kina	8.16e+01
17	45	75.0	1562	2 S3069	probable membrane pro	8.16e+01
18	44	73.3	92	2 H69749	hypothetical protein	1.22e+02
19	44	73.3	121	2 S32863	outf protein - Erwini	1.22e+02
20	44	73.3	197	2 JC4150	ribosomal protein L15	1.22e+02
21	44	73.3	230	2 S76158	hypothetical protein	1.22e+02
22	44	73.3	345	2 JG1174	amidase (EC 3.5.1.4),	1.22e+02
23	44	73.3	350	2 A69834	conserved hypotheticala	1.22e+02

24	44	73.3	401	1 WZBEB1	gene 17 protein - equ	1.22e+02
25	44	73.3	440	1 OWBY	ornithine carbamoylitr	1.22e+02
26	44	73.3	1407	2 B72078	polymorphic outer mem	1.22e+02
27	44	73.3	1610	2 A46227	voltage-dependent Ca2	1.22e+02
28	44	73.3	1646	2 JH0422	voltage-dependent cal	1.22e+02
29	44	73.3	1668	1 C69224	cobalamin biosynthesi	1.22e+02
30	44	73.3	1687	2 S41742	calcium channel alpha	1.22e+02
31	44	73.3	1851	2 T13980	calcium channel alpha	1.22e+02
32	44	73.3	1997	1 S12050	protein-tyrosine-phos	1.22e+02
33	44	73.3	2139	2 A44457	voltage-dependent cal	1.22e+02
34	44	73.3	2140	2 JH0426	voltage-dependent cal	1.22e+02
35	44	73.3	2143	2 JH0427	voltage-dependent cal	1.22e+02
36	44	73.3	2161	2 JH0564	calcium channel alpha	1.22e+02
37	44	73.3	2166	2 S11339	calcium channel prote	1.22e+02
38	44	73.3	2171	2 S05054	calcium channel alpha	1.22e+02
39	44	73.3	2181	2 A38198	calcium channel alpha	1.22e+02
40	44	73.3	2220	2 A45290	calcium channel prote	1.22e+02
41	44	73.3	2366	2 S10317	toxin B - Clostridium	1.22e+02
42	44	73.3	2367	2 S70172	photosystem II protei	1.81e+02
43	43	71.7	353	2 S42492	legumin - welwitschia	1.81e+02
44	43	71.7	519	2 S60661	semaphorin receptor V	1.81e+02
45	43	71.7	1568	2 T09074		

ALIGNMENTS

RESULT 1

ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors

H70653 #type complete
hypothetical protein Rv3843c - Mycobacterium tuberculosis
(strain H37RV)
#formal_name Mycobacterium tuberculosis
17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
H70653
A70500
Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.

#cross-references MIMD:98295987
#accession H70653
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-342 #label COL
#cross-references GB:283864; GB:AL123456; NID:g3261687; PID:e301257;
PID:gi781114
#experimental_source strain H37RV

GENETICS
#gene Rv3843C
SUMMARY
#length 342 #molecular-weight 37384 #checksum 5197

Query Match 80.0%; Score 48; DB 2; Length 342;
Best Local Similarity 50.0%; Pred. No. 2.35e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 269 WITDTQNI 276
|:|:|:
Qy 286 WITDTQNV 293

RESULT 2
ENTRY
TITLE

T09345 #type complete
hypothetical protein T26M18.70 - Arabidopsis thaliana

```

ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
          cress
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
ACCESSION T09345
REFERENCE 216650
#authors Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.;
          Duesterhoeft, A.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.;
          Lemcke, K.; Schueller, C.
#submission submitted to the Protein Sequence Database, June 1999
#accession T09345
#status Preliminary
#molecule_type DNA
##residues 1-626 #label BEV
##cross-references EMBL:AL078606
##experimental_source cultivar Columbia; BAC clone T26M18
GENETICS
#map_position 4
#introns 67/1; 106/3; 142/2; 171/3; 407/1; 484/1; 500/3; 538/3; 555/2;
          595/3
#note T26M18.70
SUMMARY #length 626 #molecular-weight 67786 #checksum 8895
Query Match 80.0%; Score 48; DB 2; Length 626;
Best Local Similarity 62.5%; Pred. No. 2.35e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 166 WIVDPQDV 173
   |||:|:|
QY 286 WITDTQNV 293

RESULT 3
ENTRY D75217 #type complete
TITLE Probable 2-acetyl-1-alkylglycerophosph ocholane esterase
PAB2176 - Pyrococcus abyssi (strain Orsay)
ORGANISM #formal_name Pyrococcus abyssi
DATE 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
ACCESSION D75217
REFERENCE A75001
#authors anonymous, Genoscope
#submission submitted to the EMBL Data Library, July 1999
#description Pyrococcus abyssi genome sequence: insights into archaeal
          chromosome structure and evolution.
#accession D75217
#status preliminary
#molecule_type DNA
##residues 1-286 #label KAW
##cross-references GB:AJ248283; GB:AL096836; NID:G5457433;
          PIDN:CA849187.1; PID:el515081; PID:G5457696
##experimental_source strain Orsay
GENETICS
#gene PAB2176
SUMMARY #length 286 #molecular-weight 32691 #checksum 5262
Query Match 76.7%; Score 46; DB 2; Length 286;
Best Local Similarity 62.5%; Pred. No. 5.42e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 257 WITDAHV 264
   ||||:|:|
QY 286 WITDTQNV 293

RESULT 4
ENTRY S33701 #type complete
TITLE homeotic protein Dth-1 - planarian (Dugesia tigrina)
ORGANISM #formal_name Dugesia tigrina
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
ACCESSION S33701; A41151
REFERENCE S33701

```

```

#authors Garcia-Fernandez, J.; Baguna, J.; Salo, E.
#journal Development (1993) 118:241-253
#title Genomic organization and expression of the planarian homeobox
          genes Dth-1 and Dth-2.
#cross-references MUID:93387216
#accession S33701
#status preliminary
#molecule_type DNA
##residues 1-533 #label GAR
##cross-references EMBL:X69200
REFERENCE A41151
#authors Garcia-Fernandez, J.; Baguna, J.; Salo, E.
#journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:7338-7342
#title Planarian homeobox genes: cloning, sequence analysis, and
          expression.
#cross-references MUID:91334461
#accession A41151
#molecule_type mRNA
##residues 83-533 #label GA2
##cross-references GB:X56499; NID:G9155; PIDN:CAA39854.1; PID:G9156
##experimental_source intact and regenerating planarians
GENETICS
#introns 135/3; 271/1
CLASSIFICATION #superfamily unassigned homeobox proteins; homeobox homology
KEYWORDS DNA binding; homeobox; nucleus; transcription regulation
FEATURE 378-434 #domain homeobox homology #label HOX
SUMMARY #length 533 #molecular-weight 60275 #checksum 7608
Query Match 76.7%; Score 46; DB 2; Length 533;
Best Local Similarity 62.5%; Pred. No. 5.42e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 279 WIKGSQNV 286
   |||:|:|
QY 286 WITDTQNV 293

RESULT 5
ENTRY A41129 #type complete
TITLE radixin - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 03-Aug-1992 #sequence_revision 14-Jul-1994 #text_change
          05-Sep-1997
ACCESSION A41129; S24201
REFERENCE A41129
#authors Funayama, N.; Nagafuchi, A.; Sato, N.; Tsukita, S.; Tsukita,
          S.
#journal J. Cell Biol. (1991) 115:1039-1048
#title Radixin is a novel member of the band 4.1 family.
#cross-references MUID:92064635
#accession A41129
#molecule_type mRNA
##residues 1-583 #label FUN
##cross-references EMBL:X60672; NID:G1033049; PID:G1334260
##note part of this sequence was confirmed by protein
          sequencing
COMMENT Radixin is a capping protein for the barbed end of actin filaments
          and it is proposed to play a role in the association of actin
          filaments with the plasma membrane.
CLASSIFICATION #superfamily ezrin; protein 4.1 membrane-binding domain
          homology
KEYWORDS actin binding; cytoskeleton
FEATURE 7-291 #domain protein 4.1 membrane-binding domain homology
          #label B41N
          470-477 #region proline-rich
          550-583 #region actin binding #status predicted
SUMMARY #length 583 #molecular-weight 68451 #checksum 1462
Query Match 76.7%; Score 46; DB 1; Length 583;
Best Local Similarity 57.1%; Pred. No. 5.42e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

Thu May 11 06:50:12 2000

Query Match 76.7%; Score 46; DB 1; Length 2204;
Best Local Similarity 57.1%; Pred. No. 5.42e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1451 WISEAQN 1457
QY 286 WITDTQN 292

RESULT 8
ENTRY TITLE
#type complete
#formal_name Aeropyrum pernix (strain K1)
#formal_name Aeropyrum pernix
#sequence_revision 20-Aug-1999 #text_change
20-Sep-1999

ACCESSIONS
REFERENCE
#authors
A72472
A72450
Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.;
Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba,
S.; Anka, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai,
Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura,
Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. (1999) 6:83-101
Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.

#cross-references MUID:99310339
#accession A72472
#status preliminary
#molecule_type DNA
#residues 1-281 #label KAW
#cross-references DDBJ:AP000064; NID:g5105945; PIDN:BAA81433.1;
#cross-references PID:d1045219; PID:g5106122
#experimental_source strain K1

GENETICS
#gene
SUMMARY
#length 281 #molecular-weight 31181 #checksum 4011
Query Match 75.0%; Score 45; DB 2; Length 281;
Best Local Similarity 50.0%; Pred. No. 8.16e+01;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 12 WVPNTQAV 19
QY 286 WITDTQN 293

RESULT 9
ENTRY TITLE
#type complete
#formal_name Caenorhabditis elegans
#sequence_revision 20-Sep-1999 #text_change
20-Sep-1999

ACCESSIONS
REFERENCE
#authors
#submission submitted to the EMBL Data Library, June 1995
#description The sequence of C. elegans cosmid F14B8.
#accession T16070
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues 1-455 #label GEI
#cross-references EMBL:U028737; NID:g860717; PID:g860719;
#cross-references PID:AAA68272.1; CESP:F14B8.3
#experimental_source strain Bristol N2

GENETICS
#gene
#introns
SUMMARY
#length 455 #molecular-weight 51505 #checksum 6201
Query Match 75.0%; Score 45; DB 2; Length 455;

RESULT 6
ENTRY TITLE
#type complete
#formal_name Synchocystis sp. (strain ICC
6803)
#formal_name Synchocystis sp.
#sequence_revision 25-Apr-1997 #text_change
21-Aug-1998

ACCESSIONS
REFERENCE
#authors
S75724
S74322
Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
DNA Res. (1996) 3:109-136
Sequence analysis of the genome of the unicellular
cyanobacterium Synchocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.

#cross-references MUID:97061201
#accession S75724
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-1014 #label KAN
#cross-references EMBL:D64003; GB:AB001339; NID:g1001200; PID:d1011110;
#cross-references PID:g1001217
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996

SUMMARY
#length 1014 #molecular-weight 115383 #checksum 9081
Query Match 76.7%; Score 46; DB 2; Length 1014;
Best Local Similarity 62.5%; Pred. No. 5.42e+01;
Matches 5; Conservative 1; Indels 0; Gaps 0;

Db 506 WIMDDQNI 513
QY 286 WITDTQN 293

RESULT 7.
ENTRY TITLE
#type complete
#formal_name Newcastle disease virus (strain
Beaudette C)
#formal_name RNA polymerase (EC 2.7.7.48)
#sequence_revision 31-Mar-1989 #text_change
11-Jun-1999

ACCESSIONS
REFERENCE
#authors
#journal
#title
#description Nucleotide sequence analysis of the L gene of Newcastle
disease virus: homologies with Sendai and vesicular
stomatitis viruses.
#cross-references MUID:87230982
#accession A26747
#molecule_type mRNA
#residues 1-2204 #label YUS
#cross-references GB:X05399; NID:g60937; PIDN:CAA28985.1; PID:g60939

GENETICS
#gene
#superfamily parainfluenza virus RNA-directed RNA polymerase
CLASSIFICATION
#keywords
SUMMARY
#length 2204 #molecular-weight 248822 #checksum 7364

#description: A directed DNA sequencing strategy based upon Tn3 transposon mutagenesis: application to the ADE1 locus on *Saccharomyces*

```

#accession S20124 chromosome I.
#molecule_type DNA
#residues 865-899, 'KDV', 903-974 #label DAV
#cross-references EMBL:M67445; NID:g170997; PIDN:AAA34400.1;
PID:g171001

REFERENCE
#authors Schweitzer, B.; Philippsen, P.
#journal Mol. Gen. Genet. (1992) 234:164-167
#title NPK1, a nonessential protein kinase gene in Saccharomyces cerevisiae with similarity to Aspergillus nidulans nima.
#cross-references MUID:92357012
#accession S25678
#molecule_type DNA
#residues 941-974 #label SCH2
#cross-references EMBL:X60549

REFERENCE
#authors Clark, M.W.; Keng, T.; Storms, R.K.; Zhong, W.; Fortin, N.; Zeng, B.; Delaney, S.; Ouellette, B.F.F.; Barton, A.B.; Kaback, D.B.; Bussey, H.
#journal Yeast (1994) 10:535-541
#title Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of the 42 kbp SPO7-CEN1-CDCL15 region.
#cross-references MUID:95028152
#accession S43452
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-974 #label CLW
#cross-references EMBL:L22015; NID:g1339990; PIDN:AAC04965.1; PID:g349757
#note the nucleotide sequence was submitted to the EMBL Data Library, November 1993

GENETICS
#gene SGD:CDCL5
#cross-references SGD:S0000072; MIPS:YAR019c
#map_position 1R
CLASSIFICATION
#superfamily unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology
#ATP; serine/threonine-specific protein kinase
#domain protein kinase homology #label KIN
#region protein kinase ATP-binding motif
#length 974 #molecular-weight 110283 #checksum 6194

Query Match 75.0%; Score 45; DB 2; Length 974;
Best Local Similarity 62.5%; Pred. No. 8.16e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 271 WINSTENV 278
QY 286 WITDTQNV 293

RESULT 13
ENTRY S54067 #type complete
TITLE Probable membrane protein YPR042c - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein YP3085.06c; hypothetical protein YP9499.01c
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 08-Jul-1995 #sequence_revision 26-Jul-1996 #text_change 12-Dec-1997
ACCESSIONS S54067; S61063
REFERENCE S54059
#authors Badcock, K.; Churcher, C.M.
#submission submitted to the EMBL Data Library, May 1995
#accession S54067
#molecule_type DNA
#residues 1-508 #label BAD
#cross-references EMBL:Z49219; NID:g805025; PID:g805026; MIPS:YPR042c
#experimental_source strain AB972
REFERENCE S61058
#authors Badcock, K.; Churcher, C.M.

```

```

#submission submitted to the EMBL Data Library, November 1995
#accession S61063
#molecule_type DNA
#residues 465-1075 #label BAW
#cross-references EMBL:Z68111; MIPS:YPR042c

GENETICS
#map_position 16R
#transmembrane protein

KEYWORDS
FEATURE 711-727 #domain transmembrane #status predicted #label TMM
SUMMARY #length 1075 #molecular-weight 119507 #checksum 3888

Query Match 75.0%; Score 45; DB 2; Length 1075;
Best Local Similarity 62.5%; Pred. No. 8.16e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 296 WITANQKV 303
QY 286 WITDTQNV 293

RESULT 14
ENTRY I48161 #type complete
TITLE p-185 precursor - golden hamster
ORGANISM #formal_name Mesocricetus auratus #common_name golden hamster
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
ACCESSIONS I48161
REFERENCE I48161
#authors Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa, T.
#journal Gene (1994) 140:251-255
#title Cloning and activation of the Syrian hamster neu proto-oncogene.
#cross-references MUID:94193007
#accession I48161
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-1254 #label RES
#cross-references GB:D16295; NID:g493236; PIDN:BA03801.1; PID:g747595

GENETICS
#gene neu
CLASSIFICATION #superfamily epidermal growth factor receptor; protein kinase homology
#ATP
KEYWORDS
FEATURE 718-983 #domain protein kinase homology #label KIN
726-734 #region protein kinase ATP-binding motif
SUMMARY #length 1254 #molecular-weight 138252 #checksum 3508

Query Match 75.0%; Score 45; DB 2; Length 1254;
Best Local Similarity 62.5%; Pred. No. 8.16e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 739 WIPDGENV 746
QY 286 WITDTQNV 293

RESULT 15
ENTRY A24571 #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
ALTERNATE_NAMES c-erb-B-2 protein precursor; kinase-related transforming protein erbB2; v-erbB-related protein HER-2/neu
ORGANISM #formal_name Homo sapiens #common_name man
DATE 25-Oct-1987 #sequence_revision 06-Dec-1996 #text_change 11-Jun-1999
ACCESSIONS A24571; A5491; A44188; B44188; I59509; I57622
REFERENCE A24571
#authors Yamamoto, T.; Ikawa, S.; Akiyama, T.; Senba, K.; Nomura, N.; Miyajima, N.; Saito, T.; Toyoshima, K.
#journal Nature (1986) 319:230-234

```

```

#title      Similarity of protein encoded by the human c-erb-B-2 gene to
#cross-references GB:86118663
#accession  A24571
#molecule_type mRNA
#residues 1-1255 #label YAM
#cross-references GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
REFERENCE
#authors    Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
#journal    Proc. Natl. Acad. Sci. U.S.A. (1985) 82:6497-6501
#title      A v-erbB-related protooncogene, c-erbB-2, is distinct from
the c-erbB-1/epidermal growth factor-receptor gene and is
amplified in a human salivary adenocarcinoma.
#cross-references MUID:86016729
#accession  A25491
#molecule_type DNA
#residues 737-1031 #label SEM
#cross-references GB:M11767; NID:g182163; PIDN:AAA35809.1; PID:g553282
REFERENCE
#authors    Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray,
A.; McGrath, J.; Seeburg, P.H.; Libermann, T.A.;
Schlessinger, J.; Francke, U.; Levinson, A.; Ullrich, A.
#journal    Science (1985) 230:1132-1139
#title      Tyrosine kinase receptor with extensive homology to EGF
receptor shares chromosomal location with neu oncogene.
#cross-references MUID:86070181
#accession  A44188
#molecule_type DNA
#residues 740-910 #label COU1
#cross-references GB:M2036; NID:g183988; PIDN:AAA35978.1; PID:g183989
#accession  B41188
#molecule_type mRNA
#residues 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255
#cross-references GB:M11730; NID:g183986
REFERENCE
#authors    King, C.R.; Kraus, M.H.; Aaronson, S.A.
#journal    Science (1985) 229:974-976
#title      Amplification of a novel v-erbB-related gene in a human
mammary carcinoma.
#cross-references MUID:85272597
#accession  I59509
#status     translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 832-909 #label REX
#cross-references GB:L29335; NID:g459807; PIDN:AAA35809.1; PID:g459808
REFERENCE
#authors    Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger,
J.; Givol, D.
#journal    Mol. Cell. Biol. (1987) 7:2597-2601
#title      Human HER2 (neu) promoter: evidence for multiple mechanisms
for transcriptional initiation.
#cross-references MUID:87286898
#accession  I57622
#status     translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-191 #label TAL
#cross-references GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
COMMENT
Amplification and overexpression of this erbB-related gene occurs
in about 30% of human breast and ovarian cancers.
GENETICS
#gene       GDB:ERBB2; NGL; NEU; HER-2
#map_position 17q21.1-17q21.1
#introns    25/1; 75/3; 147/1; 883/3
#note       the list of introns is incomplete
FUNCTION
#description catalyzes the phosphorylation of a peptidyl tyrosine residue
by ATP
CLASSIFICATION
#superfamily epidermal growth factor receptor; protein kinase
homology
KEYWORDS
ATP; autophosphorylation; duplication; glycoprotein;
phosphoprotein; phosphotransferase; proto-oncogene;

```

```

receptor; transforming protein; transmembrane protein;
tyrosine-specific protein kinase
#domain signal sequence #status predicted #label SIG\
#product protein-tyrosine kinase erbB2 #status predicted
#label MAR\
#domain extracellular #status predicted #label EXT\
#domain EGF receptor extracellular domain repeat #label
EEL\
#domain EGF receptor extracellular domain repeat #label
ER2\
#domain transmembrane #status predicted #label TM\
#domain intracellular #status predicted #label INT\
#domain protein kinase homology #label KIN\
#region protein kinase ATP-binding motif\
68,124,187,259,530,
571,629
#binding_site carbohydrate (Asn) (covalent) #status
predicted\
#binding_site phosphate (Thr) (covalent) (by protein
kinase C) #status predicted\
#active_site Lys #status predicted\
1139,1221,1222,
1248
#binding_site phosphate (Tyr) (covalent) (by
autophosphorylation) #status predicted
#length 1255 #molecular-weight 137909 #checksum 9382
SUMMARY
Query Match 75.0%; Score 45; DB 1; Length 1255;
Best Local Similarity 52.5%; Pred. No. 8.16e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 739 WIPDGENV 746
|||:|
Qy 286 WITDQNV 293
Search completed: Wed May 10 13:56:59 2000
Job time : 8 secs.

```

M I S R L A

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 13:50:32 2000; Maspar time 90.50 Seconds
2.692 Million cell updates/sec
Tubular output not generated.

Title: >US-09-376-430-2
Description: (286-293) from US09376430A.pep (21 of 25)
Perfect Score: 60
Sequence: 1 WITDTQNV 8

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 23.034; Variance 27.990; scale 0.823

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description	Pred. No.
1	46	76.7	1 Y65L_SYN7	YCF65-LIKE PROTEIN.	2.15e+01
2	46	76.7	1 PEXE_HUMAN	PEROXISOMAL MEMBRANE P	2.15e+01
3	46	76.7	1 HMH1_DUGTI	HOMEOBOX PROTEIN DPH-1	2.15e+01
4	46	76.7	1 RADI_MOUSE	RADIXIN.	2.15e+01
5	46	76.7	1 RRPL_NDV8	RNA POLYMERASE BETA SU	2.15e+01
6	45	75.0	1 YJ79_MCTU	HYPOTHETICAL 51.1 KD T	3.35e+01
7	45	75.0	1 PSQ2_YEAST	DNA CROSS-LINK REPAIR	3.35e+01
8	45	75.0	1 CC15_YEAST	CELL DIVISION CONTROL	3.35e+01
9	45	75.0	1 ERB2_MESAU	ERBB-2 RECEPTOR PROTEI	3.35e+01
10	45	75.0	1 ERB2_HUMAN	RECEPTOR PROTEIN-TYROS	3.35e+01
11	45	75.0	1 ERB2_RAT	ERBB-2 RECEPTOR PROTEI	3.35e+01
12	45	75.0	1 YMB1_YEAST	HYPOTHETICAL 180.2 KD	3.35e+01
13	44	73.3	1 GSPI_ERWCA	GENERAL SECRETION PATH	5.20e+01
14	44	73.3	1 R15E_THEAC	SOS RIBOSOMAL PROTEIN	5.20e+01
15	44	73.3	1 Y951_SYN3	HYPOTHETICAL 24.6 KD P	5.20e+01
16	44	73.3	1 OTC_YEAST	ORNITHINE CARBAMOYLTRA	5.20e+01
17	44	73.3	1 ALAM_BHOER	ALIPHATIC AMIDASE (EC	5.20e+01
18	44	73.3	1 UL43_HSVB	GENE 17 MEMBRANE PROTE	5.20e+01
19	44	73.3	1 CCAD_MESAU	VOLTAGE-DEPENDENT L-TY	5.20e+01
20	44	73.3	1 CCAM_MUSDO	MUSCLE CALCIUM CHANNEL	5.20e+01
21	44	73.3	1 CCAA_DROME	VOLTAGE-DEPENDENT CALC	5.20e+01
22	44	73.3	1 PTPB_HUMAN	PROTEIN-TYROSINE PHOSP	5.20e+01
23	44	73.3	1 CCAC_MOUSE	VOLTAGE-DEPENDENT L-TY	5.20e+01

24	44	73.3	2161	1 CCAD_HUMAN	VOLTAGE-DEPENDENT L-TY	5.20e+01
25	44	73.3	2169	1 CCAC_RAT	VOLTAGE-DEPENDENT L-TY	5.20e+01
26	44	73.3	2171	1 CCAC_RABIT	VOLTAGE-DEPENDENT L-TY	5.20e+01
27	44	73.3	2190	1 CCAD_CHICK	VOLTAGE-DEPENDENT L-TY	5.20e+01
28	44	73.3	2203	1 CCAD_RAT	VOLTAGE-DEPENDENT L-TY	5.20e+01
29	44	73.3	2221	1 CCAC_HUMAN	VOLTAGE-DEPENDENT L-TY	5.20e+01
30	44	73.3	2266	1 TOXB_CLODI	TOXIN B.	5.20e+01
31	44	73.3	2516	1 CCAD_DROME	VOLTAGE-DEPENDENT CALC	5.20e+01
32	43	71.7	352	1 PSBA_CHLEL	PHOTOSYSTEM Q(B) PROTE	7.98e+01
33	43	71.7	352	1 PSBA_PEA	PHOTOSYSTEM Q(B) PROTE	7.98e+01
34	43	71.7	353	1 PSBA_MEDSA	PHOTOSYSTEM Q(B) PROTE	7.98e+01
35	43	71.7	353	1 PSBA_HORVU	PHOTOSYSTEM Q(B) PROTE	7.98e+01
36	43	71.7	353	1 PSBA_ORYSA	PHOTOSYSTEM Q(B) PROTE	7.98e+01
37	43	71.7	353	1 PSBA_MAIZE	PHOTOSYSTEM Q(B) PROTE	7.98e+01
38	43	71.7	353	1 PSBA_CHLMO	PHOTOSYSTEM Q(B) PROTE	7.98e+01
39	43	71.7	360	1 PSB2_SINEN	PHOTOSYSTEM Q(B) PROTE	7.98e+01
40	43	71.7	360	1 PSBA_ODOSI	PHOTOSYSTEM Q(B) PROTE	7.98e+01
41	43	71.7	360	1 PSBA_ANAAZ	PHOTOSYSTEM Q(B) PROTE	7.98e+01
42	43	71.7	360	1 PSB1_ANASP	PHOTOSYSTEM Q(B) PROTE	7.98e+01
43	43	71.7	501	1 GLPD_ECOLI	AEROBIC GLYCEROL-3-PHO	7.98e+01
44	43	71.7	856	1 POLG_PVYO	GENOME POLYPROTEIN [CO	7.98e+01
45	43	71.7	856	1 POLG_PVYC	GENOME POLYPROTEIN [CO	7.98e+01

ALIGNMENTS

RESULT 1
ID Y65L_SYN7 STANDARD: PRT; 112 AA.
AC O05161;
DT 13-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE YCF65-LIKE PROTEIN.
GN MUT3G.
OS Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
RN [1]
RP SEQUENCE FROM N.A.
RA Ronen-Tarazi M., Kaplan A.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE YCF65 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U62737; AAB50399.1; -
KW Hypothetical protein.
SQ SEQUENCE 112 AA; 12764 MW; 00C732ED6CEB9287 CRC64;
Query Match 76.7%; Score 46; DB 1; Length 112;
Best Local Similarity 50.0%; Pred. No. 2.15e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db	67	WITEAERV 74	:	1
QY	286	WITDTQNV 293		
RESULT 2				
ID PEXE_HUMAN		STANDARD; PRT; 377 AA.		
AC O75381;				
DT 15-FEB-2000 (Rel. 39, Created)				
DT 15-FEB-2000 (Rel. 39, Last sequence update)				
DT 15-FEB-2000 (Rel. 39, Last annotation update)				
DE PEROXISOMAL MEMBRANE PROTEIN PEX14 (PEROXIN-14) (PEROXISOMAL MEMBRANE ANCHOR PROTEIN PEX14) (PTSI RECEPTOR DOCKING PROTEIN).				
GN PEX14.				
OS Homo sapiens (Human).				

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 98318615.
 RA "Identification of a human PTSL receptor docking protein directly
 RT required for peroxisomal protein import.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:8087-8092(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Fujiki Y.;
 RT "The peroxine Pex14p cloned by functional complementation on a CHO
 RT cell mutant mediates import of PTSL- and PTS2-proteins by interacting
 RT with Pex5p and Pex7p; dual function of a longer isoform of Pex5p.";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: COMPONENT OF THE PEROXISOMAL TRANSLLOCATION MACHINERY
 CC WITH PEX13 AND PEX17. INTERACTS WITH BOTH THE PTS1 AND PTS2
 CC RECEPTORS. BINDS DIRECTLY TO PEX17.
 CC -!- SUBCELLULAR LOCATION: PEROXISOMAL MEMBRANE-ASSOCIATED. OUTER FACE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF045186; AAC39843.1; -;
 CC EMBL; AB017546; BAA36837.1; -;
 CC MIM; 601791; -;
 KW Peroxisome; Membrane.
 SQ SEQUENCE 377 AA; 41236 MW; FED28F62A0B94E7F CRC64;

Query Match 76.7%; Score 46; DB 1; Length 377;
 Best Local Similarity 50.0%; Pred. No. 2.15e+01;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 202 WILESONI 209
 QY 286 WITDTQNV 293

RESULT 3
 ID HM1 DUCTI STANDARD; PRT; 533 AA.
 AC Q00400;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE HOMEBOX PROTEIN DTH-1.
 GN DTH-1.
 OS Dugesia tigrina (Planarian).
 CC Eukaryota; Metazoa; Platyhelminthes; Turbellaria; Tricladida;
 CC Paludicola; Dugesidae; Girardia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93387216.
 RA Garcia-Fernandez J., Baguna J., Salo E.;
 RT "Genomic organization and expression of the planarian homeobox genes
 RT Dth-1 and Dth-2.";
 RL Development 118:241-253(1993).
 RN [2]
 RP SEQUENCE OF 83-533 FROM N.A.
 RX MEDLINE; 91334461.
 RA Salo E., Garcia-Fernandez J., Baguna J.;
 RT "Planarian homeobox genes: cloning, sequence analysis, and
 RT expression.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:7338-7342(1991).
 CC -!- FUNCTION: THIS PROTEIN MIGHT BE INVOLVED IN DETERMINATION AND/OR
 CC DIFFERENTIATION OF NERVE CELLS IN THE CONTINUOUS REPLACEMENT OF
 CC NEURONS IN THE CEPHALIC REGION.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).

CC -!- TISSUE SPECIFICITY: INTESTINE AND UNIDENTIFIED PERIPHERAL
 CC PARENCHYMAL CELLS. SLIGHTLY HIGHER LEVELS IN THE CEPHALIC REGION
 CC COMPARED TO OTHER BODY REGIONS.
 CC -!- SIMILARITY: BELONGS TO THE NK-2 FAMILY OF HOMEBOX PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X69203; CAA49141.1; -;
 CC EMBL; X69200; CAA49141.1; JOINED.
 CC EMBL; X69201; CAA49141.1; JOINED.
 CC EMBL; X56499; CAA39854.1; -;
 CC PIR; A41151; A41151.
 CC PIR; S33701; S33701.
 CC HSP; P22808; IVND.
 CC PRINTS; PR00024; HOMEBOX.
 CC PROSITE; PS00027; HOMEBOX_1; 1.
 CC PROSITE; PS50071; HOMEBOX_2; 1.
 CC FRAM; PF00046; homeobox; 1.
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
 FT DOMAIN 85 90 ASN/SER-RICH.
 FT DOMAIN 100 270 PRO-RICH.
 FT DOMAIN 283 371 ASN/SER-RICH.
 FT DOMAIN 342 351 ASP/GLU-RICH (ACIDIC).
 FT DNA_BIND 377 436 HOMEBOX.
 FT DOMAIN 510 533 ASN/SER-RICH.
 SQ SEQUENCE 533 AA; 60275 MW; 2AC65D21BDEA4CAE CRC64;
 Query Match 76.7%; Score 46; DB 1; Length 533;
 Best Local Similarity 62.5%; Pred. No. 2.15e+01;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Db 279 WIKSQNV 286
 QY 286 WITDTQNV 293

RESULT 4
 ID RAD1 MOUSE STANDARD; PRT; 583 AA.
 AC P26043;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE RADIXIN.
 GN RDX.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-53 AND 263-277.
 RC TISSUE=LIVER;
 RX MEDLINE; 92064635.
 RA Funayama N., Nagafuchi A., Sato N., Tsukita S., Tsukita S.;
 RT "Radixin is a novel member of the band 4.1 family.";
 RL J. Cell Biol. 115:1039-1048(1991).
 CC -!- FUNCTION: PROBABLY PLAYS A CRUCIAL ROLE IN THE BINDING OF THE
 CC BARBED END OF ACTIN FILAMENTS TO THE PLASMA MEMBRANE.
 CC -!- SUBCELLULAR LOCATION: HIGHLY CONCENTRATED IN THE UNDERCOAT OF THE
 CC CELL-TO-CELL ADHERENS JUNCTION AND THE CLEAVAGE FURROW IN THE
 CC INTERPHASE AND MITOTIC PHASE, RESPECTIVELY.
 CC -!- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EZRIN, MOESIN,
 CC RADIXIN, AND TALIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL: X60672; CAA43087.1; -
DR PIR: A41129; A41129.
DR MGD: MGI:97887; RD.
DR PRINTS: PRO0661; ERMFAMILY.
DR PRINTS: PRO0935; BAND41.
DR PROSITE: PS00660; BAND_41_1; 1.
DR PROSITE: PS00661; BAND_41_2; 1.
DR PROSITE: PS00507; BAND_41_3; 1.
DR PFAM: PF00373; BAND_41; 1.
DR PFAM: PF00769; ERM; 1.
DR Structural protein; Cytoskeleton; Actin-binding; Capping protein.
KW DOMAIN 58 225 BAND 4.1-LIKE DOMAIN.
FT DOMAIN 470 477 POLY-PRO.
SQ SEQUENCE 583 AA; 68451 MW; 3219A52ECD87EBE CRC64;

Query Match 76.7%; Score 46; DB 1; Length 583;
Best Local Similarity 57.1%; Pred. No. 2.15e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 514 WVIETQK 520

QY 286 WITDTQN 292

RESULT 5 STANDARD; PRT; 2204 AA.
ID RRPLNDVB
AC P11205;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
DE (L PROTEIN).
GN L.

OS Newcastle disease virus (strain Beaudette C/45) (NDV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE; 8730982.
RA Yusoff K., Millar N.S., Chambers P., Emmerson P.T.;
RT "Nucleotide sequence analysis of the L gene of Newcastle disease virus: homologies with Sendai and vesicular stomatitis viruses."; Nucleic Acids Res. 15:3961-3976(1987).
RL Nucleic Acids Res. 15:3961-3976(1987).
CC -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A) SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAS, RNA EDITING OF THE P GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL: X05399; CAA28985.1; -
DR PIR: A26747; RRMZNV.
DR PFAM: PF00946; Paramyx_RNA_pol; 1.
KW Transferase; RNA-directed RNA polymerase.
SQ SEQUENCE : 2204 AA; 248822 MW; C67B86674D904802C CRC64;

Query Match 76.7%; Score 46; DB 1; Length 2204;
Best Local Similarity 57.1%; Pred. No. 2.15e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1451 WISEAQN 1457

QY 286 WITDTQN 292

RESULT 6 STANDARD; PRT; 481 AA.
ID VJ79 MYCTU
AC Q10875; O53980;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 51.1 KD TRANSPORT PROTEIN RV1979C.
GN RV1979C OR MTCY39.40 OR MTV051.17C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37RV;
MEDLINE: 98295987.
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaita F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."; Nature 393:537-544(1998).
RL Nature 393:537-544(1998).
CC -!- FUNCTION: PROBABLE AMINO-ACID OR METABOLITE TRANSPORT PROTEIN.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC -----
EMBL: AL022073; CAA17852.1; -
DR PROSITE; PS00218; AMINO_ACID_PERMEASE; FALSE_NEG.
DR TUBERCULIST; RV1979C; -
KW Hypothetical protein; Transport; Transmembrane.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 46 66 POTENTIAL.
FT TRANSMEM 90 110 POTENTIAL.
FT TRANSMEM 134 154 POTENTIAL.
FT TRANSMEM 167 187 POTENTIAL.
FT TRANSMEM 218 238 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 303 323 POTENTIAL.
FT TRANSMEM 377 397 POTENTIAL.
FT TRANSMEM 411 431 POTENTIAL.
FT TRANSMEM 446 466 POTENTIAL.
SQ SEQUENCE 481 AA; 51102 MW; 4E5F5E90DC32B9D1 CRC64;

Query Match 75.0%; Score 45; DB 1; Length 481;
Best Local Similarity 62.5%; Pred. No. 3.35e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 183 WITTHQV 190

QY 286 WITDTQN 293

RESULT 7

ID PS02_YEAST STANDARD; PRT; 661 AA.
AC P30620;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA CROSS-LINK REPAIR PROTEIN PSO2/SNM1.

GN PS02 OR SNM1 OR YMR137C OR YMR9375.06C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE; 92140357.
 RA Richter D., Niegemann E., Brendel M.;
 RT "Molecular structure of the DNA cross-link repair gene SNM1 (PS02) of
 RT the yeast Saccharomyces cerevisiae";
 RL Mol. Gen. Genet. 231:194-200(1992).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN-S288C / AB972;
 RC Baddock K., Churcher C., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN DNA CROSS-LINK REPAIR.
 CC -1- SIMILARITY: MAY CONTAIN ONE ZINC-BINDING DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X64004; CAA45405.1; -
 DR EMBL; 247071; CAA87351.1; -
 DR PIR; S19646; S19646.
 DR SGD; L0001953; PS02.
 KW DNA repair; Zinc-finger.
 FT ZN_FING 145 169
 SQ SEQUENCE 661 AA; 76398 MW; 56F14DBAC86EAE2 CRC64;

 Query Match 75.0%; Score 45; DB 1; Length 661;
 Best Local Similarity 75.0%; Pred. No. 3.35e+01;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 304 WITDRISV 311
 QY 286 WITDTQNV 293

 RESULT 8
 ID CC15_YEAST STANDARD; PRT; 974 AA.
 AC P27636;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE CELL DIVISION CONTROL PROTEIN 15 (EC 2.7.1.-).
 GN CDC15 OR YAR019C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE; 92357012.
 RA Schweitzer B., Philippson P.;
 RT "NPK1, a nonessential protein kinase gene in Saccharomyces cerevisiae
 RT with similarity to Aspergillus nidulans nima.";
 RL Mol. Gen. Genet. 234:164-167(1992).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN-GRF88;
 RX MEDLINE; 91353080.
 RA Schweitzer B., Philippson P.;
 RT "CDC15, an essential cell cycle gene in Saccharomyces cerevisiae,
 RT encodes a protein kinase domain.";
 RL Yeast 7:265-273(1991).
 RN [3]
 RP REVISIONS TO 900-902.
 RA Schweitzer B.;

Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.
 [4]
 RN SEQUENCE FROM N.A.
 RP STRAIN-S288C / AB972;
 RX MEDLINE; 95028152.
 RA Clark M.W., Keng T., Storms R.K., Zhong W., Fortin N., Zeng B.,
 RA Delaney S., Ouellette B.F.F., Barton A.B., Kaback D.B., Bussey H.;
 RT "Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of
 RT the 42 kbp SPO7-CEN1-CDC15 region";
 RL Yeast 10:535-541(1994).
 RN [5]
 RN SEQUENCE OF 865-974 FROM N.A.
 RP Davies C.J., Hutchison C.A. III;
 RA Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: ESSENTIAL FOR LATE NUCLEAR DIVISION IN THE MITOTIC
 CC CYCLE. REQUIRED FOR INACTIVATION OF CDC2 KINASE ACTIVITY AT THE
 CC END OF MITOSIS.
 CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
 CC PROTEIN KINASES.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-4 IS THE INITIATOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X60549; CAA43041.1; -
 DR EMBL; X52683; CAA36906.1; -
 DR EMBL; L22015; AAC04965.1; ALT_SEQ.
 DR EMBL; M67445; AAA34400.1; -
 DR PIR; S15038; S15038.
 DR PIR; S25680; S25680.
 DR HSP; P24941; IAQ1.
 DR SGD; L0000255; CDC15.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PFAM; PF00069; pkinase; 1.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;
 KW Cell cycle; Cell division; Mitosis.
 FT DOMAIN 25 272
 FT NP_BIND 31 39 ATP (BY SIMILARITY).
 FT BINDING 54 54 ATP (BY SIMILARITY).
 FT ACT_SITE 146 146 BY SIMILARITY.
 FT ACT_SITE 146 146 BY SIMILARITY.
 SQ SEQUENCE 974 AA; 110352 MW; 7A69AB7FEE291991 CRC64;

 Query Match 75.0%; Score 45; DB 1; Length 974;
 Best Local Similarity 62.5%; Pred. No. 3.35e+01;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Db 271 WINSTENV 278
 QY 286 WITDTQNV 293

 RESULT 9
 ID ERB2_MESAU STANDARD; PRT; 1254 AA.
 AC Q60553;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE ERBB-2 RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112)
 DE (P185ERBB2) (NEU PROTO-ONCOGENE).
 GN ERBB2 OR NEU.
 OS Mesocricetus auratus (Golden hamster).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.
 [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE-NERVE;

```

RX MEDLINE; 94193007.
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -!- FUNCTION: NEUREGULINS AND GP30 ARE POTENTIAL LIGANDS FOR THIS
CC RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIREGULIN.
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D16295; BAA03801.1; -
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC PFAM; PF00069; pkinase; 1.
CC PFAM; PF00757; Furin-like; 1.
CC PFAM; PF01030; Recep_L_domain; 2.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21
FT CHAIN 22 1254 ERBB-2 RECEPTOR PROTEIN-TYROSINE KINASE.
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 653 675 POTENTIAL.
FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 158 368 CYS-RICH.
FT DOMAIN 472 644 CYS-RICH.
FT DOMAIN 720 987 PROTEIN KINASE.
FT NP_BIND 726 734 ATP (BY SIMILARITY).
FT BINDING 753 753 ATP (BY SIMILARITY).
FT ACT_SITE 845 845 BY SIMILARITY.
FT VARIANT 658 658 V -> E (IN ONCOGENIC NEU).
FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 POTENTIAL.
FT CARBOHYD 125 125 POTENTIAL.
FT CARBOHYD 187 187 POTENTIAL.
FT CARBOHYD 259 259 POTENTIAL.
FT CARBOHYD 530 530 POTENTIAL.
FT CARBOHYD 571 571 POTENTIAL.
FT CARBOHYD 629 629 POTENTIAL.
SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 75.0%; Score 45; DB 1; Length 1254;
Best Local Similarity 62.5%; Pred. No. 3.35e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 739 WIPDGENV 746
|||::|
QY 286 WITDQNV 293

RESULT 10
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)

```

```

DE RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2 PRECURSOR (EC 2.7.1.112)
DE (P185ERBB2) (NEU PROTO-ONCOGENE) (C-ERBB-2) (TYROSINE KINASE-TYPE CELL
DE SURFACE RECEPTOR HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86118663.
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erb-B-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85070181.
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Franke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE; 86016729.
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE; 93194196.
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993).
CC -!- FUNCTION: NEUREGULINS AND GP30 ARE POTENTIAL LIGANDS FOR THIS
CC RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIREGULIN.
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -!- POLYMORPHISM: THERE ARE FOURS ALLELES DUE TO THE VARIATIONS IN
CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M11767; AAA35808.1; -
CC EMBL: M11761; AAA35808.1; JOINED.
CC EMBL: M11762; AAA35808.1; JOINED.
CC EMBL: M11763; AAA35808.1; JOINED.
CC EMBL: M11764; AAA35808.1; JOINED.
CC EMBL: M11765; AAA35808.1; JOINED.
CC EMBL: M11766; AAA35808.1; JOINED.
CC EMBL: M11730; AAA75493.1; -
CC EMBL: M12036; AAA35978.1; -
CC EMBL: X03363; CAA27060.1; -
CC FIR; A25491; A25491.

```

DR PIR; A24571; A24571.
 DR HSP; P11362; 1FGI.
 DR MIM; 164870; .
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PFAM; PF00069; pkinase; 1.
 DR PFAM; PF00757; Furin-like; 1.
 DR PFAM; PF01030; Recep_L-domain; 2.
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Polymorphism. 1 21
 FT SIGNAL 1 21
 FT CHAIN 22 1255
 FT DOMAIN 22 652
 FT TRANSMEM 653 675
 FT DOMAIN 676 1255
 FT DOMAIN 720 987
 FT NP_BIND 726 734
 FT BINDING 753 753
 FT ACT_SITE 845 845
 FT MOD_RES 1139 1139
 FT MOD_RES 1248 1248
 FT CARBOHYD 68 68
 FT CARBOHYD 124 124
 FT CARBOHYD 187 187
 FT CARBOHYD 259 259
 FT CARBOHYD 530 530
 FT CARBOHYD 571 571
 FT CARBOHYD 629 629
 FT VARIANT 654 654
 FT VARIANT 655 655
 FT CONFLICT 1170 1170
 FT SEQUENCE 1255 AA; 137909 MW; 39E9DFDA04DC962 CRC64;
 Query Match 75.0%; Score 45; DB 1; Length 1255;
 Best Local Similarity 62.5%; Pred. No. 3.35e+01;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Db 739 WIPDGENV 746
 QY 286 WITDTQNV 293
 RESULT 11
 ID ERB2_RAT STANDARD; PRT; 1257 AA.
 AC P06494;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ERBB-2 RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112)
 DE (P185ERBB2) (NEU PROTO-ONCOGENE) (EPIDERMAL GROWTH FACTOR RECEPTOR-RELATED PROTEIN).
 DE ERBB2 OR NEU.
 GN Rattus norvegicus (Rat).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-NEUROBLASTOMA;
 RX MEDLINE; 86118662.
 RA Bargmann C.I.; Hung M.-C.; Weinberg R.A.;
 RT "The neu oncogene encodes an epidermal growth factor receptor-related protein."
 RL Nature 319:226-230(1986).
 [2]
 RP SEQUENCE OF 852-905 FROM N.A.
 RC TISSUE-SCIATIC NERVE;
 RX MEDLINE; 91222560.
 RA Lai C.; Lemke G.;
 RT "An extended family of protein-tyrosine kinase genes differentially

RT expressed in the vertebrate nervous system.";
 RL Neuron 6:691-704(1991).
 RN [3]
 RP STRUCTURE BY NMR OF 650-666.
 RX MEDLINE; 92155181.
 RA Gullick W.J.; Bottomley A.C.; Lofts F.J.; Doak D.G.; Mulvey D.;
 RA Newman R.; Crumpton M.J.; Sternberg M.J.E.; Campbell I.D.;
 RT "Three dimensional structure of the transmembrane region of the proto-oncogenic and oncogenic forms of the neu protein.";
 RL EMBO J. 11:43-48(1992).
 CC -1- FUNCTION: NEUREGULINS AND GP30 ARE POTENTIAL LIGANDS FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIREGULIN.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP + PROTEIN TYROSINE PHOSPHATE.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS. THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 DR EMBL; X03362; CAA27059.1; ALT_INIT.
 DR PIR; A24562; TVRTNU.
 DR HSP; P11362; 1FGI.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR PFAM; PF00069; pkinase; 1.
 DR PFAM; PF00757; Furin-like; 1.
 DR PFAM; PF01030; Recep_L-domain; 2.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21
 FT CHAIN 22 1257
 FT DOMAIN 22 654
 FT TRANSMEM 655 677
 FT DOMAIN 678 1257
 FT DOMAIN 159 369
 FT DOMAIN 473 646
 FT DOMAIN 722 989
 FT NP_BIND 728 736
 FT BINDING 755 755
 FT ACT_SITE 847 847
 FT VARIANT 661 661
 FT MOD_RES 1141 1141
 FT MOD_RES 1250 1250
 FT CARBOHYD 68 68
 FT CARBOHYD 188 188
 FT CARBOHYD 260 260
 FT CARBOHYD 532 532
 FT CARBOHYD 573 573
 FT CARBOHYD 631 631
 SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;
 Query Match 75.0%; Score 45; DB 1; Length 1257;
 Best Local Similarity 62.5%; Pred. No. 3.35e+01;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Db 741 WIPDGENV 748
 QY 286 WITDTQNV 293

RESULT 12

```
ID YM81_YEAST STANDARD; PRT; 1562 AA.
AC Q04781; Q04029;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-NOV-1997 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 180.2 KD PROTEIN IN FAA4-HOR7 INTERGENIC REGION.
GN YMR247C OR YMR408.09C OR YMR920.01C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE OF 1-956 FROM N.A.
RC STRAIN=S288C / AB972;
RA Hunt S., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 950-1562 FROM N.A.
RC STRAIN=S288C / AB972;
RA Gentles S., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO HUMAN KIA00714.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z48639; CAA88574.1; -.
CC EMBL; Z48756; CAA88657.1; -.
CC HYPOTHETICAL protein.
CC KW Hypothetical protein.
CC QY SEQUENCE 1562 AA; 180185 MW; 97AC65E881362305 CRC64;
CC
Query Match 75.0%; Score 45; DB 1; Length 1562;
Best Local Similarity 62.5%; Pred. No. 3.35e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 1472 WIMSTQHV 1479
QY 286 WITDTQNV 293
|||
RESULT 13
ID GPI_ERWCA STANDARD; PRT; 121 AA.
AC P31588;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE GENERAL SECRETION PATHWAY PROTEIN I PRECURSOR (PECTIC ENZYMES
DE SECRETION PROTEIN OUT1).
GN OUT1.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Erwinia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCRI193;
RX MEDLINE; 93316842.
RA Bunce N., Barallion R., Douglas P., Mulholland V., Stevens S.,
RA Walker S., Salmond G.P.C.;
RT "Molecular cloning and characterization of 13 out genes from Erwinia
RT carotovora subspecies carotovora: genes encoding members of a general
RT secretion pathway (GSP) widespread in gram-negative bacteria.";
RL Mol. Microbiol. 8:443-456(1993).
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLLOCATION OF THE
CC MULTIPLE PECTIC ENZYMES.
CC -!- SIMILARITY: BELONGS TO THE PULI/OUT1/XPSI/EXEI/XCPV FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X70049; CAA49650.1; -.
CC EMBL; S31752; S31752.
CC EMBL; S32863; S32863.
CC PROSITE; PS00409; PROKAR_NTER_METHYL; 1.
KW Transport; Methylation.
FT PROPEP 1 6 BY SIMILARITY.
FT CHAIN 7 121 GENERAL SECRETION PATHWAY PROTEIN I.
FT MOD_RES 7 7 METHYLATION (BY SIMILARITY).
FT SEQUENCE 121 AA; 13813 MW; FE32FD2D8A5C1627 CRC64;
CC
Query Match 73.3%; Score 44; DB 1; Length 121;
Best Local Similarity 37.5%; Pred. No. 5.20e+01;
Matches 3; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 48 WVAENQOV 55
QY 286 WITDTQNV 293
|||||
RESULT 14
ID R15E_THEAC STANDARD; PRT; 197 AA.
AC P49403;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE 50S RIBOSOMAL PROTEIN L15E.
GN RPL15E OR RPL15.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;
OC Thermoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95251669.
RA Zwickl P., Lupas A., Baumeister W.;
RT "The Thermoplasma acidophilum rpl15 gene encodes a homologue of
RT eukaryotic ribosomal proteins L15/YL10.";
RL Biochem. Biophys. Res. Commun. 209:684-688(1995).
CC -!- SIMILARITY: BELONGS TO THE L15E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U16148; AAA68967.1; -.
CC PROSITE; PS01194; RIBOSOMAL_L15E; 1.
CC PRAM; PF00827; Ribosomal_L15e; 1.
CC KW Ribosomal protein.
CC QY SEQUENCE 197 AA; 22928 MW; 589910E54174F848 CRC64;
CC
Query Match 73.3%; Score 44; DB 1; Length 197;
Best Local Similarity 57.1%; Pred. No. 5.20e+01;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 151 WISEPQN 157
QY 286 WITDTQNV 292
|||||
RESULT 15
ID Y951_SYNY3 STANDARD; PRT; 230 AA.
AC P74323;
DT 01-NOV-1997 (Rel. 35, Created)
```

DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DI 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL 24.6 KD PROTEIN SLR0951.
 GN SLR0951.
 OS Synchocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97061201.
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
 RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
 RA Yamada M., Yasuda M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -!- SIMILARITY: BELONGS TO THE UPF0007 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D90914; BAA18417.1; -
 DR PROSITE; PS01295; UPF0007; 1.
 DR PFAM; PF01128; UPF0007; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 230 AA; 24643 MW; D5E5AD5893E98E92 CRC64;

 Query Match 73.3%; Score 44; DB 1; Length 230;
 Best Local Similarity 71.4%; Pred. No. 5.20e+01;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 Db 145 WITDTPD 151
 ||||| :
 Qy 286 WITDTQN 292

Search completed: Wed May 10 13:52:11 2000
 Job time : 99 secs.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Wed May 10 13:52:30 2000; MasPar time 230.97 Seconds
Tabular output not generated. 2.401 Million cell updates/sec

Title: >US-09-376-430-2
Description: (286-293) from US09376430A.pep (21 of 25)
Perfect Score: 60
Sequence: 1 WITDQNV 8

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 22.528; Variance 26.815; scale 0.840

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description	Pred. No.
1	49	81.7	039734	POLYPROTEIN.	8.73e+00
2	48	80.0	342 2	HYPOTHETICAL 37.4 KD P	1.42e+01
3	47	78.3	565 10	(EC 3.2.2.22) (RRNA N-	2.29e+01
4	47	78.3	566 10	(EC 3.2.2.22) (RRNA N-	2.29e+01
5	47	78.3	568 5	E02H4.2 PROTEIN.	2.29e+01
6	47	78.3	2035	MYELOBLAST KIAA0233.	2.29e+01
7	46	76.7	155 14	HYPOTHETICAL 18.0 KD P	3.66e+01
8	46	76.7	376 11	PEROXISOMAL MEMBRANE A	3.66e+01
9	46	76.7	377 4	PEROXISOMAL MEMBRANE A	3.66e+01
10	46	76.7	377 11	PEROXISOMAL MEMBRANE A	3.66e+01
11	46	76.7	398 5	HYPOTHETICAL 42.3 KD P	3.66e+01
12	46	76.7	1014 2	HYPOTHETICAL 115.4 KD	3.66e+01
13	46	76.7	2204 14	LARGE PROTEIN.	3.66e+01
14	46	76.7	2204 14	LARGE POLYMERASE PROTE	3.66e+01
15	45	75.0	197 4	CYTOPLASMIC DYNEIN 3 H	5.81e+01
16	45	75.0	216 3	SNM 1-2 TS (FRAGMENT).	5.81e+01
17	45	75.0	281 1	281AA LONG HYPOTHETICA	5.81e+01
18	45	75.0	455 5	SIMILARITY WITH VARIOU	5.81e+01
19	45	75.0	558 10	ZETA-CAROTENE DESATUR	5.81e+01
20	45	75.0	700 5	STIALIDASE (EC 3.2.1.18	5.81e+01

21	45	75.0	1075	3	Q12221	HYPOTHETICAL 119.5 KD	5.81e+01
22	45	75.0	1259	6	O18735	ERBB-2.	5.81e+01
23	44	73.3	92	2	O31454	YBEN PROTEIN.	9.16e+01
24	44	73.3	101	3	Q00029	PEPTIDE SYNTHETASE (FR	9.16e+01
25	44	73.3	146	14	Q87014	(SCSV7).	9.16e+01
26	44	73.3	220	2	O51814	HYPOTHETICAL 25.1 KD P	9.16e+01
27	44	73.3	294	5	O18517	CALCIUM CHANNEL ALPHA-	9.16e+01
28	44	73.3	350	2	O07564	HYPOTHETICAL 39.3 KD P	9.16e+01
29	44	73.3	362	2	O92B60	NRPU.	9.16e+01
30	44	73.3	395	2	O54095	PUTATIVE OXIDASE.	9.16e+01
31	44	73.3	445	2	O9X7H2	PUTATIVE SODIUM/CHLORI	9.16e+01
32	44	73.3	567	10	O04947	ASCORATE OXIDASE PREC	9.16e+01
33	44	73.3	617	2	O56794	XANTHOMONAS COMPESTRIS	9.16e+01
34	44	73.3	649	3	O43069	HYPOTHETICAL 73.4 KD P	9.16e+01
35	44	73.3	1407	2	O92B99	POLYMORPHIC OUTER MEMB	9.16e+01
36	44	73.3	1426	11	P70298	CUT-LIKE 2 (CUX-2).	9.16e+01
37	44	73.3	1668	1	O27011	CORALAMIN BIOSYNTHESIS	9.16e+01
38	44	73.3	1783	5	O17050	PUTATIVE L-TYPE CALCIU	9.16e+01
39	44	73.3	1891	5	O97017	L-TYPE CALCIUM CHANNEL	9.16e+01
40	44	73.3	1906	5	O18698	C48A7.1 PROTEIN.	9.16e+01
41	44	73.3	2049	5	O20849	CODED FOR BY C. ELEGAN	9.16e+01
42	44	73.3	2169	11	O92305	L-TYPE CALCIUM CHANNEL	9.16e+01
43	44	73.3	2367	2	O46034	TOXIN B.	9.16e+01
44	44	73.3	3766	5	Q17551	SIMILAR TO REGULATOR O	9.16e+01
45	43	71.7	3081	14	Q85105	POLYPROTEIN (FRAGMENT)	1.43e+02

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT; 3493 AA.
ID 039734	AC 039734;	
AC 039734;	DT 01-JAN-1998 (TrEMBLrel. 05, Created)	
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)	DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)	
DE POLYPROTEIN.	OS Sweet potato feathery mottle virus.	
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;	OC Potyvirus.	
RN [1]	SEQUENCE FROM N.A.	
RC STRAIN-S STRAIN;	RA NISHIGUCHI M.;	
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.	[2]	
RN SEQUENCE FROM N.A.	RC STRAIN-S STRAIN;	
RC STRAIN-S STRAIN;	EX MEDLINE; 98336489.	
RA SAKAI J., MORI M., MORISHITA T., TANAKA M., HANADA K., USUGI T.,	RA NISHIGUCHI M.;	
RT "Complete nucleotide sequence and genome organization of sweet potato feathery mottle virus (S strain) genomic RNA: the large coding region of the P1 gene."	RT Arch. Virol. 142:1553-1562(1997).	
RL Arch. Virol. 142:1553-1562(1997).	DR EMBL; D86371; BAA22702.1; -	
DR EMBL; D86371; BAA22702.1; -	DR PFAM; PF00863; Peptidase_C4; 1.	
DR PFAM; PF00863; Peptidase_C4; 1.	DR PFAM; PF00851; Peptidase_C6; 1.	
DR PFAM; PF00767; Poty_coat; 1.	DR PFAM; PF01577; Poty_P1; 1.	
DR PFAM; PF01577; Poty_P1; 1.	DR PFAM; PF00680; RNA_dep_RNA_pol; 1.	
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.	DR PRINTS; PR00966; NIAPOTYPASE.	
DR PRINTS; PR00966; NIAPOTYPASE.	FT CHAIN 1	664
FT CHAIN 1	FT CHAIN 665	1122
FT CHAIN 665	FT CHAIN 1123	1474
FT CHAIN 1123	FT CHAIN 1475	1526
FT CHAIN 1475	FT CHAIN 1527	2169
FT CHAIN 1527	FT CHAIN 2170	2222
FT CHAIN 2170	FT CHAIN 2223	2414
FT CHAIN 2223	FT CHAIN 2415	2657
FT CHAIN 2415	FT CHAIN 2658	3178
FT CHAIN 2658	FT CHAIN 3179	3493
FT CHAIN 3179	SEQUENCE 3493 AA; A02EFA05 CRC32;	

Query Match 81.7%; Score 49; DB 14; Length 3493;
 Best Local Similarity 62.5%; Pred. No. 8.73e+00;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1185 WITDQEV 1192
 |||:|:|
 QY 286 WITDTQNV 293

RESULT 2
 ID P96235; PRELIMINARY; PRT; 342 AA.
 AC F96235;
 DT 01-MAY-1997 (TREMELrel. 03, Created)
 DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)
 DT 01-NOV-1998 (TREMELrel. 08, Last annotation update)
 DE HYPOHETICAL 37.4 KD PROTEIN.
 GN MTCY01A6.26.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RA BADCOCK K., CHURCHER C.M.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RA BARRELL B.G., RAJANDREAM M.A.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE; 96181548.
 RA PHILIPP W.J., POULET S., EIGMEIER K., PASCOPELLA L.,
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
 RA COLE S.T.;
 RT "An integrated map of the genome of the tubercle bacillus,
 RT Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium
 RT leprae.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
 DR EMBL; 283864; CAB06223.1;
 KW Hypothetical protein.
 SQ SEQUENCE 342 AA; 37384 MW; 7EC67739 CRC32;

Query Match 80.0%; Score 48; DB 2; Length 342;
 Best Local Similarity 50.0%; Pred. No. 1.42e+01;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 269 WITDAQGI 276
 |||:|:|
 QY 286 WITDTQNV 293

RESULT 3
 ID O04071; PRELIMINARY; PRT; 565 AA.
 AC O04071;
 DT 01-JUL-1997 (TREMELrel. 04, Created)
 DT 01-JUL-1997 (TREMELrel. 04, Last sequence update)
 DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)
 DE (EC 3.2.2.22) (RNA N-GLYCOSIDASE) PRECURSOR.
 GN RIP.
 OS Sambucus nigra (European elder).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Asterales; euasterids II; unclassified euasterids II;
 OC Adoxaceae; Sambucus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97236787.
 RA VAN DAMME E.J., BARRE A., ROUGE P., VAN LEUVEN F., PEUMANS W.J.;
 RT "Isolation and molecular cloning of a novel type 2 ribosome-

inactivating protein with an inactive B chain from elderberry
 (Sambucus nigra) bark.";
 J. Biol. Chem. 272:8353-8360(1997).
 CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF THE N-GLYCOSIDIC BOND AT ONE
 SPECIFIC ADENOSINE ON THE 28S RRNA.
 DR EMBL; U58357; AAC49672.1;
 DR HSSP; P02879; 2AAI.
 DR MENDEL; 15494; Samni; Rip; 15494.
 DR PROSITE; PS00275; SHIGA_RICIN; 1.
 DR PFAM; PF00652; Ricin_B_lectin; 2.
 DR PFAM; PF00161; RIP; 1.
 DR PRINTS; PR00396; SHIGARICIN.
 KW Signal; Hydrolase; Glycosidase; Toxin.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 565 RIBOSOME INACTIVATING PROTEIN, A AND B
 CHAIN.
 FT CHAIN 26 283 RIBOSOME INACTIVATING PROTEIN, A CHAIN.
 FT CHAIN 305 565 RIBOSOME INACTIVATING PROTEIN, B CHAIN.
 SQ SEQUENCE 565 AA; 62348 MW; F5BB9CA6 CRC32;

Query Match 78.3%; Score 47; DB 10; Length 565;
 Best Local Similarity 75.0%; Pred. No. 2.29e+01;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 558 WITDTPV 565
 |||||:|
 QY 286 WITDTQNV 293

RESULT 4
 ID O04072; PRELIMINARY; PRT; 566 AA.
 AC O04072;
 DT 01-JUL-1997 (TREMELrel. 04, Created)
 DT 01-JUL-1997 (TREMELrel. 04, Last sequence update)
 DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)
 DE (EC 3.2.2.22) (RNA N-GLYCOSIDASE) PRECURSOR.
 GN RIP.
 OS Sambucus nigra (European elder).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Asteridae; euasterids II; unclassified euasterids II;
 OC Adoxaceae; Sambucus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97236787.
 RA VAN DAMME E.J., BARRE A., ROUGE P., VAN LEUVEN F., PEUMANS W.J.;
 RT "Isolation and molecular cloning of a novel type 2 ribosome-
 inactivating protein with an inactive B chain from elderberry
 (Sambucus nigra) bark.";
 J. Biol. Chem. 272:8353-8360(1997).
 CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF THE N-GLYCOSIDIC BOND AT ONE
 SPECIFIC ADENOSINE ON THE 28S RRNA.
 DR EMBL; U58358; AAC49673.1;
 DR HSSP; P02879; 2AAI.
 DR MENDEL; 15495; Samni; Rip; 15495.
 DR PROSITE; PS00275; SHIGA_RICIN; 1.
 DR PFAM; PF00652; Ricin_B_lectin; 2.
 DR PFAM; PF00161; RIP; 1.
 DR PRINTS; PR00396; SHIGARICIN.
 KW Signal; Hydrolase; Glycosidase; Toxin.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 566 RIBOSOME INACTIVATING PROTEIN, A AND B
 CHAIN.
 FT CHAIN 23 284 RIBOSOME INACTIVATING PROTEIN, A CHAIN.
 FT CHAIN 305 566 RIBOSOME INACTIVATING PROTEIN, B CHAIN.
 SQ SEQUENCE 566 AA; 62733 MW; 4BB4ECB6 CRC32;

Query Match 78.3%; Score 47; DB 10; Length 566;
 Best Local Similarity 75.0%; Pred. No. 2.29e+01;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 559 WITDTPV 566
 |||||:|


```

QY 286 WITDTQNV 293

RESULT 5
ID Q19039 PRELIMINARY; PRT; 568 AA.
AC Q19039;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)
DE E02H4.2 PROTEIN.
GN E02H4.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RN SEQUENCE FROM N.A.
RA BARLOW K.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BURTON J.,
RA BONFELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z68003; CAA91976.1;
SQ SEQUENCE 568 AA; 63726 MW; 6CCEDF4D6 CRC32;

Query Match 78.3%; Score 47; DB 5; Length 568;
Best Local Similarity 71.4%; Pred. No. 2.29e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 75 WIPETQN 81
II :III
QY 286 WITDTQNV 292

RESULT 6
ID Q92508 PRELIMINARY; PRT; 2035 AA.
AC Q92508;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE MYELOBLAST KIAA0233.
GN KIAA0233.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RN SEQUENCE FROM N.A.
RA TISSUE-BONE MARROW;
RX MEDLINE; 97191544.
RA NAGASE T., SEKI N., ISHIKAWA K., OHIRA M., KAWARABAYASI Y., OHARA O.,
RA TANAKA A., KOTANI H., MIYAJIMA N., NOMURA N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
DR EMBL; D87071; BAA13240.1;
SQ SEQUENCE 2035 AA; 233040 MW; 31324A29 CRC32;

Query Match 78.3%; Score 47; DB 4; Length 2035;
Best Local Similarity 50.0%; Pred. No. 2.29e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 202 WILESONI 209
II :III
QY 286 WITDTQNV 293

Db 965 WVTNAQAV 972
II :I
QY 286 WITDTQNV 293

RESULT 7
ID Q9WIJ6 PRELIMINARY; PRT; 155 AA.
AC Q9WIJ6;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE HYPOTHETICAL 18.0 KD PROTEIN.
GN C3-EG.
OS faba bean necrotic yellows virus.
OC Viruses; ssDNA viruses; Circoviridae.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-EGYPTIAN EVI-93;
RX MEDLINE; 99094637.
RA KATUL L., TIMCHENKO T., GRONENBORN B., VETTEN H.J.;
RT "Ten distinct circular ssDNA components, four of which encode putative
RT replication-associated proteins, are associated with the faba bean
RT necrotic yellows virus genome.";
RL J. Gen. Virol. 79:3101-3109(1998).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN-EGYPTIAN EVI-93;
RA TIMCHENKO T., DE KOUCHKOVSKY F., KATUL L., DAVID C., VETTEN H.J.;
RA GRONENBORN B.;
RT "Analysis of five FBVIV-associated Rep proteins.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ132181; CAB44021.1;
KW Hypothetical protein.
SQ SEQUENCE 155 AA; 17980 MW; 219885E8 CRC32;

Query Match 76.7%; Score 46; DB 14; Length 155;
Best Local Similarity 50.0%; Pred. No. 3.66e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 129 WIVNTEDV 136
II :I
QY 286 WITDTQNV 293

RESULT 8
ID Q9Z224 PRELIMINARY; PRT; 376 AA.
AC Q9Z224;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE PEROXISOMAL MEMBRANE ANCHOR PROTEIN.
GN PEX14.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RN SEQUENCE FROM N.A.
RA FUJIKI Y.;
RT "The peroxine Pex14p cloned by functional complementation on a CHO
RT cell mutant mediates import of PTS1- and PTS2-proteins by interacting
RT with Pex5p and Pex7p: Dual function of a longer isoform of Pex5p.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017544; BAA36835.1;
SQ SEQUENCE 376 AA; 40968 MW; BF060DFB CRC32;

Query Match 76.7%; Score 46; DB 11; Length 376;
Best Local Similarity 50.0%; Pred. No. 3.66e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 202 WILESONI 209
II :III
QY 286 WITDTQNV 293

```

```

RESULT 9
ID O75381 PRELIMINARY; PRT; 377 AA.
AC O75381;
DT 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE PEROXISOMAL MEMBRANE ANCHOR PROTEIN HSPX14P.
GN PEX14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98318615.
RA FRANKEN S., TERLECKY S.R., SUBRAMANI S.;
RT "Identification of a human PEX1 receptor docking protein directly
RT required for peroxisomal protein import.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8087-8092(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA FUJIKI Y.;
RT "The peroxine Pex14p cloned by functional complementation on a CHO
RT cell mutant mediates import of PEX1- and PEX2-proteins by interacting
RT with Pex5p and Pex7p: Dual function of a longer isoform of Pex5p.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF045186; AAC39843.1; -.
DR EMBL; AB017546; BAA36837.1; -.
SQ SEQUENCE 377 AA; 41236 MW; 7128916C CRC32;

Query Match 76.7%; Score 46; DB 4; Length 377;
Best Local Similarity 50.0%; Pred. No. 3.66e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 202 WILESQNI 209
QY 286 WITDTQNV 293
II :||:

RESULT 10
ID Q92223 PRELIMINARY; PRT; 377 AA.
AC Q92223;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE PEROXISOMAL MEMBRANE ANCHOR PROTEIN.
GN PEX14.
OS Cricetulus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
RN [1]
RP SEQUENCE FROM N.A.
RA FUJIKI Y.;
RT "The peroxine Pex14p cloned by functional complementation on a CHO
RT cell mutant mediates import of PEX1- and PEX2-proteins by interacting
RT with Pex5p and Pex7p: Dual function of a longer isoform of Pex5p.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017545; BAA36836.1; -.
SQ SEQUENCE 377 AA; 41196 MW; F4A14AC4 CRC32;

Query Match 76.7%; Score 46; DB 11; Length 377;
Best Local Similarity 50.0%; Pred. No. 3.66e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 202 WILESQNI 209
QY 286 WITDTQNV 293
II :||:

RESULT 11
ID O15713 PRELIMINARY; PRT; 398 AA.
AC O15713;

```

```

DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)
DE HYPOTHETICAL 42.3 KD PROTEIN.
OS Toxoplasma gondii.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;
OC Toxoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-PLK;
RX MEDLINE; 98107666.
RA KNOLL L.J., BOOTHROYD J.C.;
RT "Isolation of developmentally regulated genes from Toxoplasma gondii
RT by a gene trap with the positive and negative selectable marker
RT hypoxanthine-xanthine-guanine phosphoribosyltransferase.";
RL Mol. Cell. Biol. 18:807-814(1998).
DR EMBL; AF015290; AAC38825.1; -.
KW Hypothetical protein.
SQ SEQUENCE 398 AA; 42345 MW; 30986A97 CRC32;

Query Match 76.7%; Score 46; DB 5; Length 398;
Best Local Similarity 57.1%; Pred. No. 3.66e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 299 WVTETQK 305
QY 286 WITDTQN 292
II :||:

RESULT 12
ID Q55374 PRELIMINARY; PRT; 1014 AA.
AC Q55374;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TRENBLrel. 09, Last annotation update)
DE HYPOTHETICAL 115.4 KD PROTEIN.
OC Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-PCC6803;
RA TABATA S.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-PCC6803;
RX MEDLINE; 96127529.
RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
RA SUGIURA M., TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAKI K., KIMURA T.,
RA HOSOUCHI T., MATSUO A., MURAKI A., NAKAZAKI N., NAROK K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D64003; BAA10459.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1014 AA; 115383 MW; 45267BCB CRC32;

Query Match 76.7%; Score 46; DB 2; Length 1014;
Best Local Similarity 62.5%; Pred. No. 3.66e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

Db 506 WIMDDQNI 513
 |||
 QY 286 WITDTQNV 293

RESULT 13
 ID Q9WH6 PRELIMINARY; PRT; 2204 AA.
 AC Q9WH6;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
 DE LARGE PROTEIN.
 GN L.
 OS Newcastle disease virus.
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA ROMER-OBEDORFER A., BUCHHOLZ U.J., MUNDT E., METTENLEITER T.C.;
 RT "Generation of recombinant lentogenic Newcastle Disease Virus from
 RT cDNA."
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA ROMER-OBEDORFER A.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Y18898; CAB51327.1; -.
 SQ SEQUENCE 2204 AA; 248500 MW; EBA13B27 CRC32;

Query Match 76.7%; Score 46; DB 14; Length 2204;
 Best Local Similarity 57.1%; Pred. No. 3.66e+01;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1451 WISEAQN 1457
 |||
 QY 286 WITDTQNV 292

RESULT 14
 ID Q90341 PRELIMINARY; PRT; 2204 AA.
 AC Q90341;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE LARGE POLYMERASE PROTEIN.
 GN L.
 OS Newcastle disease virus.
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LASOTA;
 RA DE LEEUW O.S., PEETERS B.P.H.;
 RT "Complete nucleotide sequence of Newcastle disease virus: evidence for
 RT the existence of a new genus within the subfamily Paramyxovirinae."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF077761; AAC28375.1; -.
 DR PFAM; PF00946; Paramyx_RNA_pol; 1.
 SQ SEQUENCE 2204 AA; 248675 MW; DE9FC87D CRC32;

Query Match 76.7%; Score 46; DB 14; Length 2204;
 Best Local Similarity 57.1%; Pred. No. 3.66e+01;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1451 WISEAQN 1457
 |||
 QY 286 WITDTQNV 292

RESULT 15
 ID Q92816 PRELIMINARY; PRT; 197 AA.
 AC Q92816;

DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
 DE CYTOPLASMIC DYNEIN 3 HEAVY CHAIN (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96234671.
 RA VAISBERG E.A., GRISSOM P.M., MCINTOSH J.R.;
 RT "Mammalian cells express three distinct dynein heavy chains that are
 RT localized to different cytoplasmic organelles."
 RL J. Cell Biol. 133:831-842(1996).
 DR EMBL; U53532; AAB09729.1; -.
 DR HSSP; P02649; 1LE4.. 1
 FT NON_TER 197 1
 FT NON_TER 197 1
 SQ SEQUENCE 197 AA; 22099 MW; 9B8487BB CRC32;

Query Match 75.0%; Score 45; DB 4; Length 197;
 Best Local Similarity 50.0%; Pred. No. 5.81e+01;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1 WTSETQEV 8
 |::|:
 QY 286 WITDTQNV 293

Search completed: Wed May 10 13:56:33 2000
 Job time : 243 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MFsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 13:19:00 2000; MasPar time 228.24 Seconds
Tabular output not generated. 5.772 Million cell updates/sec

Title: >US-09-376-430-2
Description: (151-169) from US09376430A.pep (17 of 25)
Sequence: 1 QYRSPFDTEWQSKQNTCN 19

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 27.853; Variance 40.348; scale 0.690

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description	Pred. No.
1	70	48.3	437	5	Q20820 F55C5.6 PROTEIN.	1.47e-01
2	68	46.9	495	2	O67163 CYSTEINYL-TRNA SYNTHET	3.51e-01
3	66	45.5	787	2	O06776 HYPOTHETICAL 86.1 KD P	8.24e-01
4	65	44.8	554	5	Q22815 T26C5.2 PROTEIN.	1.26e+00
5	64	44.1	816	10	O9X156 F9L1.7 PROTEIN.	1.91e+00
6	63	43.4	222	3	O08278 CHROMOSOME XV READING	2.88e+00
7	63	43.4	394	2	Q52962 CHEMOTAXIS MOTB PROTEI	2.88e+00
8	63	43.4	401	10	Q92Y3 CALRETICULIN PRECURSOR	2.88e+00
9	62	42.8	201	10	O23099 BAC TM018A10.	4.34e+00
10	62	42.8	479	5	O45138 R02D3.7 PROTEIN.	4.34e+00
11	62	42.8	1239	2	Q55138 HYPOTHETICAL 137.8 KD	4.34e+00
12	62	42.8	5105	5	O61201 F39C12.1 PROTEIN.	4.34e+00
13	61	42.1	584	10	O81118 CELL WALL INVERTASE (E	6.51e+00
14	60	41.4	247	14	Q9YMS8 LDORF-46 PEPTIDE.	9.71e+00
15	60	41.4	268	2	O24834 4-OXALOCROTONATE DECAR	9.71e+00
16	60	41.4	308	5	O45530 F49A5.4 PROTEIN.	9.71e+00
17	60	41.4	450	14	Q9WDF0 NEURAMINIDASE (FRAGMEN	9.71e+00
18	60	41.4	1028	11	O07409 PLASMACYTOMA-ASSOCIATE	9.71e+00
19	60	41.4	1127	5	O62205 F32B4.3B PROTEIN.	9.71e+00
20	60	41.4	1165	5	O62206 F32B4.3A PROTEIN.	9.71e+00

21	59	40.7	112	9	Q38665 NINX PROTEIN.	1.44e+01
22	59	40.7	126	2	Q32202 XVOI PROTEIN.	1.44e+01
23	59	40.7	209	3	Q23397 XYLANASE PRECURSOR.	1.44e+01
24	59	40.7	257	2	Q92KRO PUTATIVE AMINO ACID AB	1.44e+01
25	59	40.7	268	4	O95278 LAFORIN (FRAGMENT).	1.44e+01
26	59	40.7	346	1	Q48941 MOLYBDENUM-CONTAINING	1.44e+01
27	59	40.7	378	5	O44157 CD4.5 PROTEIN.	1.44e+01
28	59	40.7	490	2	O84709 GTPASE/GTP-BINDING PRO	1.44e+01
29	59	40.7	496	10	Q9ZNS8 CYTOSOLIC GLUTATHIONE	1.44e+01
30	59	40.7	529	2	O31303 OLIGOPEPTIDE PERMEASE	1.44e+01
31	59	40.7	529	2	O50927 OLIGOPEPTIDE ABC TRANS	1.44e+01
32	59	40.7	881	5	Q23613 ZK822.1 PROTEIN.	1.44e+01
33	59	40.7	890	2	O86727 PUTATIVE SECRETED CELL	1.44e+01
34	59	40.7	1028	11	P97528 NB-3	1.44e+01
35	59	40.7	1777	14	O89278 NONSTRUCTURAL PROTEIN	1.44e+01
36	58	40.0	118	11	Q9WT74 NG38.	2.13e+01
37	58	40.0	388	5	O16881 C10F3.2 PROTEIN.	2.13e+01
38	58	40.0	432	5	O61820 B0511.10 PROTEIN.	2.13e+01
39	58	40.0	625	11	Q9WV42 RFG.	2.13e+01
40	58	40.0	680	5	O01849 SIMILARITY TO RAT AND	2.13e+01
41	58	40.0	990	2	O06512 SITE-SPECIFIC RECOMBIN	2.13e+01
42	58	40.0	1026	11	Q82845 NEURAL CELL ADHESION P	2.13e+01
43	58	40.0	1614	11	O88883 LOW-DENSITY LIPOPROTEI	2.13e+01
44	58	40.0	1615	4	O75197 LIPOPROTEIN RECEPTOR	2.13e+01
45	58	40.0	2322	14	Q9WN12 POLYPROTEIN PRECURSOR.	2.13e+01

ALIGNMENTS

RESULT	ID	Q20820	PRELIMINARY;	PRT;	437 AA.
AC	Q20820;				
DT	01-NOV-1996	(TrEMBLrel. 01, Created)			
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)			
DT	01-NOV-1999	(TrEMBLrel. 12, Last annotation update)			
DE	F55C5.6	PROTEIN.			
GN	F55C5.6				
OS	Caenorhabditis elegans.				
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;				
OC	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	HARRIS B.;				
RL	Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 94150718.				
RA	WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,				
RA	BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,				
RA	CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,				
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,				
RA	JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,				
RA	LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,				
RA	PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,				
RA	SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,				
RA	THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,				
RT	WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;				
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."				
RL	Nature 368:32-38(1994).				
DR	EMBL; 278198; CAB01571.1;				
DR	PFAM; PF01593; Amino_Oxidase; 1.				
SQ	SEQUENCE 437 AA; 50263 MW; 3AC0F120 CRC32;				

Query Match 48.3%; Score 70; DB 5; Length 437;
Best Local Similarity 35.3%; Pred. No. 1.47e-01;
Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
Db 178 YKNOFOIWSASAEIC 194
QY 152 YRSPFDTEWQSKQNTCN 168.

RESULT 2
ID O67163 PRELIMINARY; PRT; 495 AA.
AC O67163;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE CYSSTEINYL-TRNA SYNTHETASE.
GN CYSS.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE; 98196666.
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus";
RL Nature 392:353-358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE000721; AAC07125.1;
DR PFAM; PF01406; trna-synC_1e; 1.
DR PRINTS; PR00983; TRNASYNTHCS.
KW Aminoacyl-trna synthetase.
SQ SEQUENCE 495 AA; 57135 MW; E84F4020 CRC32;

Query Match 46.9%; Score 68; DB 2; Length 495;
Best Local Similarity 46.7%; Pred. No. 3.51e-01;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
Db 297 HYRSPDLFSWKEE 311
QY 151 QYRSPFDTEWQKQE 165
:||||:| : :|

RESULT 3
ID O06776 PRELIMINARY; PRT; 787 AA.
AC O06776;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 86.1 KD PROTEIN.
GN MT1376.13C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA MURPHY L., HARRIS D.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA BARRELL B.G., RAJANDREAM M.A., PARKHILL J.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA COLE S.T.;
RT "An integrated map of the genome of the tubercle bacillus,
Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium
leprae.";

RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR EMBL; Z95972; CAB09383.1; -;
DR PFAM; PF00884; Sulfatase; 1.
KW Hypothetical protein.
SQ SEQUENCE 787 AA; 86127 MW; 5E38F5FF CRC32;
Query Match 45.5%; Score 66; DB 2; Length 787;
Best Local Similarity 43.8%; Pred. No. 8.24e-01;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
Db 256 YRGKFDVGVNDALREET 271
QY 152 YRSPFDTEWQKQENT 167
||: || :| :|

RESULT 4
ID Q22815 PRELIMINARY; PRT; 554 AA.
AC Q22815;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)
DE T26C5.2 PROTEIN.
GN T26C5.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA THOMAS K.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROCK J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z50859; CAA90727.1; -;
SQ SEQUENCE 554 AA; 63085 MW; A838A92B CRC32;

Query Match 44.8%; Score 65; DB 5; Length 554;
Best Local Similarity 42.9%; Pred. No. 1.26e+00;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
Db 38 PYHSEWDMKTEDCMC 51
QY 155 PFDTWQSKQENTC 168
|: ||: | :|

RESULT 5
ID Q9X156 PRELIMINARY; PRT; 816 AA.
AC Q9X156;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 14, Last annotation update)
DE F9L1.7 PROTEIN.
GN F9L1.7.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eukaryophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;
 RA VYSOTSKAYA V.S., SCHWARTZ J.R., YU G., TORIUMI M., LENZ C., LIU S.,
 RA LEE J., LIU A., LI J., KREMENTSKAYA I., LUROS J., GONZALEZ A.,
 RA ALTAFI H., ARAUJO R., BROOKS S., BUEHLER E., CHAO O., CONN L.,
 RA CONWAY A.B., DUNN P., HANSEN N., HUIZAR L., KHAN S., KIM C., PALM C.,
 RA ROWLEY D., SHINN P., WALKER M., DAVIS R.W., ECKER J.R.,
 RA FEDERSPIEL N.A., THEOLOGIS A.,
 RT "Arabidopsis thaliana chromosome 1 BAC F9L1 sequence."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC007591; AAD39642.1; -
 SQ SEQUENCE 816 AA; 91812 MW; 36C989BB CRC32;

Query Match 44.1%; Score 64; DB 10; Length 816;
 Best Local Similarity 47.1%; Pred. No. 1.91e+00;
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 Db 449 QFRSFGTRWRPOSST 465
 QY 151 QYRSPFDTEWQSK 167

RESULT 6
 ID O08278 PRELIMINARY; PRT; 222 AA.
 AC O08278;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE CHROMOSOME XV READING FRAME ORF YOL135C
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA ARINO J., CASAMAYOR A., GAMO F.J., GANCEDO C., LAFUENTE M.J.,
 RA ALDEA M., CASAS C., HERRERO E.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MTPS;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FY1679;
 RX MEDLINE; 97051593.
 RA ALDEA M., PIEDRAFITA L., CASAS C., CASAMAYOR A., KHALID H.,
 RA BALCELLS L., ARINO J., HERRERO E.;
 RT "Sequence analysis of a 12 801 bp fragment of the left arm of yeast
 RT chromosome XV containing a putative 6-phosphofructo-2-kinase gene, a
 RT gene for a possible glycopospholipid-anchored surface protein and six
 RT other open reading frames."
 RL Yeast 12:1053-1058(1996).
 DR EMBL: 274877; CAA99156.1; -
 DR EMBL: X95465; CAA64734.1; -
 SQ SEQUENCE 222 AA; 25585 MW; 7AA4AF9E CRC32;

Query Match 43.4%; Score 63; DB 3; Length 222;
 Best Local Similarity 46.7%; Pred. No. 2.88e+00;
 Matches 7; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
 Db 75 QYRA-FGSIWQVKDQ 88
 QY 151 QYRSPFDTEWQSK 165

RESULT 7
 ID Q52962 PRELIMINARY; PRT; 394 AA.
 AC Q52962;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE CHEMOTAXIS MOTB PROTEIN (MOTILITY PROTEIN B).
 GN MOTB.
 OS Rhizobium meliloti.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RUI1/001;
 RA PLATZER J., STERR W., HAUSMANN M., SCHMITT R.;
 RL J. Bacteriol. 179:6391-6399(1997).
 CC -!- FUNCTION: REQUIRED FOR THE ROTATION OF THE FLAGELLAR MOTOR. MIGHT
 CC BE A LINKER THAT FASTENS THE TORQUE-GENERATING MACHINERY TO THE
 CC CELL WALL.
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -!- SIMILARITY: TO MOTB IN OTHER BACTERIA.
 DR EMBL: L49337; AAB81408.1; -
 DR PFAM: PF00691; OmpA; 1.
 KW Chemotaxis; Flagella; Transmembrane; Flagellar rotation.
 FT DOMAIN 1 32 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 33 52 POTENTIAL.
 FT DOMAIN 53 394 EXTRACELLULAR (POTENTIAL).
 SQ SEQUENCE 394 AA; 41963 MW; FD2BD5DA CRC32;

Query Match 43.4%; Score 63; DB 2; Length 394;
 Best Local Similarity 57.1%; Pred. No. 2.88e+00;
 Matches 8; Conservative 2; Mismatches 2; Indels 2; Gaps 2;
 Db 176 YRDPFDPDFW-TKQ 188
 QY 152 YRSPFDTEWQSK 164

RESULT 8
 ID Q92NV3 PRELIMINARY; PRT; 401 AA.
 AC Q92NV3;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE CALRETICULIN PRECURSOR.
 OS Euglena gracilis.
 OC Eukaryota; Eulenzozoa; Euglenida; Euglenales; Euglena.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA NAVAZZIO L., BALDAN B., MARTIN W., MARIANI P.;
 RT "Evidence for conservation of a calcium homeostat component:
 RT purification characterization and cloning of calreticulin from Euglena
 RT gracilis."
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: 109816; CAA70945.1; -
 FT CHAIN 19 401 CALRETICULIN.
 SQ SEQUENCE 401 AA; 45910 MW; FA62AA60 CRC32;

Query Match 43.4%; Score 63; DB 10; Length 401;
 Best Local Similarity 25.0%; Pred. No. 2.88e+00;
 Matches 3; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
 Db 22 YKTFEPDWETR 33
 QY 152 YRSPFDTEWQSK 163

RESULT 9
 ID O23099 PRELIMINARY; PRT; 201 AA.
 AC O23099;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
 DE BAC TM018A10.
 GN A-TM018A10.11.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;
RA DEMPSEY S., HARPER M.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA WATERSTON R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF013294; AAB62856.1;
SQ SEQUENCE 201 AA; 23876 MW; A0283412 CRC32;

Query Match 42.8%; Score 62; DB 10; Length 201;
Best Local Similarity 50.0%; Pred. No. 4.34e+00;
Matches 8; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

Db 14 SPFNLDGWSMDEASC 29
QY 154 SPFD-TWQSKOENTC 168

RESULT 10
ID O44505 PRELIMINARY; PRT; 479 AA.
AC O44505;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE R02D3.7 PROTEIN.
GN R02D3.7
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]

Query Match 42.8%; Score 62; DB 5; Length 479;
Best Local Similarity 47.4%; Pred. No. 4.34e+00;
Matches 9; Conservative 3; Mismatches 5; Indels 2; Gaps 2;

Db 226 QSRDQF-KWFFQREDTC 243
QY 151 QYRSPDTEW-QSKOENTC 168

RESULT 11
ID Q55138 PRELIMINARY; PRT; 1239 AA.

AC Q55138;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE HYPOTHETICAL 137.8 KD PROTEIN.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RX MEDLINE; 96127529.
RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
RA SUGIURA M., TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome."
RL DNA Res. 2:153-166(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
DR EMBL; D64001; BAAL0273.1;+.
SQ SEQUENCE 1239 AA; 137824 MW; 129C451F CRC32;

Query Match 42.8%; Score 62; DB 2; Length 1239;
Best Local Similarity 72.7%; Pred. No. 4.34e+00;

Matches 8; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db 480 QYRI-FDPEWQ 489
QY 151 QYRSPDTEWQ 161

RESULT 12
ID O61201 PRELIMINARY; PRT; 5105 AA.

AC O61201;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE F39C12.1 PROTEIN.
GN F39C12.1

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA CHISSOE S., SANSONE J.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF039043; AAB94194.1;
 SQ SEQUENCE 5105 AA; 585903 MW; 850C353C CRC32;

Query Match 42.8%; Score 62; DB 5; Length 5105;
 Best Local Similarity 41.7%; Pred. No. 4.34e+00;
 Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 3689 SDWOKKNDKCN 3700
 :||| : :
 Qy 158 TEWQSKQENTCN 169

RESULT 13 PRELIMINARY; PRT; 584 AA.
 ID O81118
 AC O81118;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
 DE CELL WALL INVERTASE (EC 3.2.1.26).
 GN IVRI.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euryhalophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 OC Poaceae; Triticum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-KATEPWA;
 RA MINHAS J.S., SAINI H.S.;
 RL "Invited genes from wheat anthers.";
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 CC -I- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
 DR EMBL; AF030420; AAC96085.1;
 DR MENDEL; 31769; Triae;1002;31769.
 DR PROSITE; PS00609; GLYCOSYL-HYDROL_F32; 1.
 DR PFAM; PF00251; Glyco_hydro_32; 1.
 KW Hydrolase; Glycosidase.
 SQ SEQUENCE 584 AA; 66240 MW; 66A4A603 CRC32;

Query Match 42.1%; Score 61; DB 10; Length 584;
 Best Local Similarity 42.9%; Pred. No. 6.51e+00;
 Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 425 EDTAWQDAQGLCK 438
 :||| : :
 Qy 156 FDEWQSKQENTCN 169

RESULT 14 PRELIMINARY; PRT; 247 AA.
 ID Q9YMS8
 AC Q9YMS8;
 DT 01-MAY-1999 (TReMBLrel. 10, Created)
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
 DE LCORF-46 PEPTIDE.
 OS Lymantria dispar multicapsid nuclear polyhedrosis virus (LdMNPV).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolyhedrovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE; 99124785.

RA KUZIO J., PEARSON M.N., HARWOOD S.H., FUNK C.J., EVANS J.T.,
 RA SLAVICEK J.M., ROHRMANN G.F.;
 RT "Sequence and analysis of the genome of a baculovirus pathogenic for
 RT Lymantria dispar.";
 RL Virology 253:17-34(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA KUZIO J., PEARSON M.N., HARWOOD S.H., FUNK C.J., EVANS J.T.,
 RA SLAVICEK J., ROHRMANN G.F.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA KUZIO J.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF081810; AAC70231.1;
 SQ SEQUENCE 247 AA; 28494 MW; 82DEEE0 CRC32;

Query Match 41.4%; Score 60; DB 14; Length 247;
 Best Local Similarity 60.0%; Pred. No. 9.71e+00;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

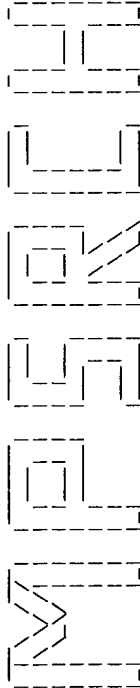
Db 80 YRAPFVLHWQ 89
 :||| : :
 Qy 152 YRSPFDTEWQ 161

RESULT 15 PRELIMINARY; PRT; 268 AA.
 ID O24834
 AC O24834;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
 DE 4-OXALOCROTONATE DECARBOXYLASE.
 GN ATDH.
 OS Acinetobacter sp.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
 OC Moraxellaceae; Acinetobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YAA;
 RX MEDLINE; 9717776.
 RA FUJII T., TAKEO M., MAEDA Y.;
 RT "Plasmid-encoded genes specifying aniline oxidation from Acinetobacter
 RT sp. strain YAA.";
 RL Microbiology 143:93-99(1997).
 DR EMBL; AB008831; BAA23561.1;
 SQ SEQUENCE 268 AA; 29356 MW; 27685740 CRC32;

Query Match 41.4%; Score 60; DB 2; Length 268;
 Best Local Similarity 30.8%; Pred. No. 9.71e+00;
 Matches 4; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 43 YKDAFDIQWEIRR 55
 : :||| : : :
 Qy 152 YRSPFDTEWQSKQ 164

Search completed: Wed May 10 13:22:59 2000
 Job time : 239 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsearch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 13:32:04 2000; MasPar time 3.15 Seconds
Tabular output not generated. 142.655 Million cell updates/sec

Title: >US-09-376-430-2
Description: (188-206) from US09376430A.pep (18 of 25)
Perfect Score: 144
Sequence: 1 KAMEDVYGPDTYPSDWSEV 19

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:geneseqp

Statistics: Mean 21.156; Variance 72.822; scale 0.291

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	64	44.4	846	1 R74691	Human very low density	4.74e+01
2	64	44.4	846	1 R74692	Rat very low density 1	4.74e+01
3	64	44.4	873	1 W02212	Human VLDL receptor.	4.74e+01
4	64	44.4	873	1 R44735	apo-E lipoprotein rece	4.74e+01
5	61	42.4	151	1 Y11141	S. pneumoniae peptide	8.99e+01
6	61	42.4	207	1 W38549	S. pneumoniae peptide	8.99e+01
7	61	42.4	229	1 W32621	Bacillus smithii nitri	8.99e+01
8	61	42.4	435	1 Y04390	Human origin of recogn	8.99e+01
9	61	42.4	514	1 W33359	Streptococcus pneumoni	8.99e+01
10	60	41.7	77	1 W97675	Mitogen activated prot	1.11e+02
11	60	41.7	219	1 W78195	Human secreted protein	1.11e+02
12	60	41.7	1006	1 R07115	Gene product of yeast	1.11e+02
13	60	41.7	1007	1 W17518	Aspergillus niger beta	1.11e+02
14	59	41.0	159	1 R07564	Polyepitide encoded sa	1.37e+02
15	58	40.3	115	1 W28023	Anaerobic ribonucleosi	1.69e+02
16	58	40.3	377	1 W05411	Human H74 protein.	1.69e+02
17	58	40.3	552	1 W48873	Hyphozyma sp. strain C	1.69e+02
18	58	40.3	573	1 W48874	Hyphozyma sp. strain C	1.69e+02
19	58	40.3	925	1 R79148	Human insulin receptor	1.69e+02
20	57	39.6	501	1 R74676	Arabidopsis thaliana a	2.08e+02
21	57	39.6	537	1 W68463	Bacillus agaradhaerens	2.08e+02
22	57	39.6	911	1 R32508	A. oryzae beta-galacto	2.08e+02
23	57	39.6	1005	1 W11238	Beta-galactosidase.	2.08e+02

24	56	38.9	234	1 W82694	P. neopolydactyla type	2.55e+02
25	56	38.9	286	1 W64058	Human IL-9 receptor pr	2.55e+02
26	56	38.9	387	1 P70502	Human protein C inhibi	2.55e+02
27	56	38.9	500	1 W64057	Human IL-9 receptor pr	2.55e+02
28	56	38.9	501	1 W64056	Human IL-9 receptor pr	2.55e+02
29	56	38.9	501	1 W64055	Human IL-9 receptor pr	2.55e+02
30	56	38.9	518	1 R78680	Tobacco leaf antifunga	2.55e+02
31	56	38.9	540	1 R78679	Rabbit growth hormone	2.55e+02
32	56	38.9	638	1 P92107	Rabbit growth hormone	2.55e+02
33	56	38.9	638	1 P81327	Rabbit growth hormone	2.55e+02
34	56	38.9	638	1 W33395	Bovine enterokinase.	2.55e+02
35	56	38.9	798	1 R57283	Mycoplasma pneumoniae	2.55e+02
36	56	38.9	1030	1 W19602	Human FRAG1 protein.	3.13e+02
37	55	38.2	122	1 W41593	Sequence of pre-S(1) (3.13e+02
38	55	38.2	122	1 P60283	C. albicans Cpr prote	3.13e+02
39	55	38.2	168	1 W64387	Human interleukin-4 re	3.13e+02
40	55	38.2	197	1 W50026	Human TR6 protein.	3.13e+02
41	55	38.2	411	1 W76827	Derived amino acid seq	3.13e+02
42	55	38.2	824	1 R04575	Human IL-4 receptor cl	3.13e+02
43	55	38.2	825	1 W93004	Human IL-4 receptor.	3.13e+02
44	55	38.2	825	1 R11254	IL4-R-GBP 130 fusion p	3.13e+02
45	55	38.2	918	1 R70120		

ALIGNMENTS

RESULT 1
ID R74691 standard; Protein; 846 AA.
AC R74691;
DT 21-JAN-1996 (first entry)
DE Human very low density lipoprotein receptor.
KW VLDL receptor; very low density lipoprotein receptor;
KW hyperlipidaemia; cardiovascular disease; disease diagnosis;
KW atherosclerosis; hypercholesterolemia.
OS Homo sapiens.
PN WO9513374-A2.
PD 18-MAY-1995.
PF 08-NOV-1994; U12911.
PR 08-NOV-1993; US-149103.
PA (BAYU) BAYLOR COLLEGE MEDICINE.
PI Chan LCB.
DR WPI; 95-194093/25.
DR N-PSDB; Q88687.
PT Nucleic acid encoding very low density lipoprotein receptor - used to develop prods. for treating e.g. hyperlipidaemia for screening assays and for diagnostic imaging
PS Claim 2; Page 42; 59pp; English.
CC This protein may be expressed recombinantly in a transgenic animal. By elevating levels of a VLDL receptor in an animal, the receptors will aid in removal of circulating VLDL and related lipoproteins to decrease the risk of developing coronary diseases. It may be used in the treatment of e.g. hyperlipidaemia, atherosclerosis or hypercholesterolemia.
SQ Sequence 846 AA;

Query Match 44.4%; Score 64; DB 1; Length 846;
Best Local Similarity: 37.5%; Pred. No. 4.74e+01;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 480 KMDNVNPAIAVDW 495
QY 188 KAMEDVYGPDTYPSDW 203

RESULT 2
ID R74692 standard; Protein; 846 AA.
AC R74692;

DT 21-JAN-1996 (first entry)
DE Rat very low density lipoprotein receptor.
KW VLDL receptor; very low density lipoprotein receptor;
KW hyperlipidaemia; cardiovascular disease; disease diagnosis;
KW atherosclerosis; hypercholesterolemia.
OS Rattus rattus.

PN W09513374-A2.
 PD 18-MAY-1995.
 PF 08-NOV-1994; U12911.
 PR 08-NOV-1993; US-149103.
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 PI Chan LCB;
 DR WPI; 95-194093/25.
 DR N-PSDB; Q88688.
 PT Nucleic acid encoding very low density lipoprotein receptor - used
 PT to develop prods. for treating e.g. hyperlipidaemia for screening
 PT assays and for diagnostic imaging
 PS Claim 2; Page 44; 59pp; English.
 CC This protein may be expressed recombinantly in a transgenic animal.
 CC By elevating levels of a VLDL receptor in an animal, the receptors
 CC will aid in removal of circulating VLDL and related lipoproteins to
 CC decrease the risk of developing coronary diseases. It may be used
 CC in the treatment of e.g. hyperlipidaemia, atherosclerosis or
 CC hypercholesterolemia.
 CC Sequence 846 AA;

Query Match 44.4%; Score 64; DB 1; Length 846;
 Best Local Similarity 37.5%; Pred. No. 4.74e+01;
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 480 KMIIDVYNPAIAVDW 495
 QY 188 KAMEDVYGPDTYPSDW 203

RESULT 3

ID W02212 standard; Protein; 873 AA.

AC W02212;

DT 09-NOV-1996 (first entry)

DE Human VLDL receptor.

KW Very low density lipoprotein receptor; VLDL; adenovirus; vector;

KW gene therapy; familial hypercholesterolemia; hyperlipidaemia;

KW hepatocyte; liver.

OS Homo sapiens.

PN W09626286-A1.

PD 29-AUG-1996.

PF 23-FEB-1996; U03041.

PR 24-FEB-1995; US-393734.

PA (UYPE) UNIV PENNSYLVANIA.

PI Kozarsky K, Strauss J, Wilson JM;

DR WPI; 96-402376/40.

DR N-PSDB; T36751.

PT Recombinant adenoviral vector contg. human VLDL receptor gene -

PT providing stable expression in liver cells, for gene therapy of

PT familial hyper-cholesterolemia and combined hyper-lipidaemia

PS Disclosure; Page 66-69; 106pp; English.

CC The amino acid sequence (W02212) of human very low density

CC lipoprotein (VLDL) receptor has been deduced from a previously

CC isolated cDNA clone (T36751) (Gafvels et al., Som. Cell. Mol.

CC Genet., 19:557-569, 1993). The receptor has been suggested

CC to play a role in the uptake of triglyceride-rich lipoprotein

CC particles by specific organs. A new use for the cDNA clone

CC involves its incorporation into an adenoviral vector (see also

CC T36752), and delivery to hepatocytes for the gene therapy of

CC familial hypercholesterolemia and familial combined hyperlipidaemia.

CC Sequence 873 AA;

Query Match 44.4%; Score 64; DB 1; Length 873;
 Best Local Similarity 37.5%; Pred. No. 4.74e+01;
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 507 KMIIDVYNPAIAVDW 522

QY 188 KAMEDVYGPDTYPSDW 203

RESULT 4

ID R44735 standard; Protein; 873 AA.

AC R44735;

DT 06-JUN-1994 (first entry)
 DE apo-E lipoprotein receptor.
 KW apo-E lipoprotein receptor; tissue specificity; LDL;
 KW low density lipoprotein; spleen; heart; binding activity.
 OS Oryctolagus cuniculus.
 PN J05294998-A.
 PD 09-NOV-1993.
 PF 13-APR-1992; 092638.
 PR 13-APR-1992; JP-092638.
 PA (SANY) SANKYO CO LTD.
 DR WPI; 93-392866/49.
 DR N-PSDB; Q53137.

PT Rabbit apo-E lipoprotein receptor - with DNA sequence coding
 PT specified aminoacid in protein
 PS Claim 1; Page 7; 14pp; Japanese.
 CC The sequence shows an apo-E lipoprotein receptor. The gene was
 CC isolated from a cDNA library extracted from liver, heart and spleen
 CC of a normal rabbit. The gene was cloned and expressed in G418-
 CC resistant cells. Positive clones were isolated by testing their
 CC beta-very LDL binding activity. Transfected cells showed LDL and
 CC VLDL binding activity. The receptor has high tissue specificity.
 CC Sequence 873 AA;

Query Match 44.4%; Score 64; DB 1; Length 873;
 Best Local Similarity 37.5%; Pred. No. 4.74e+01;
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 507 KMIIDVYNPAIAVDW 522

QY 188 KAMEDVYGPDTYPSDW 203

RESULT 5

ID Y11141 standard; Protein; 151 AA.

AC Y11141;

DT 20-MAY-1999 (first entry)

DE S. pneumoniae peptide chain release factor 3.

KW Streptococcus pneumoniae strain 0100993; vaccine; immune response;

KW streptococcal infection; pneumococcal.

OS Streptococcus pneumoniae.

PN W09737026-A1.

PD 09-OCT-1997.

PF 01-APR-1997; U05306.

PR 22-AUG-1996; US-025788.

PR 02-APR-1996; US-014690.

PA (SMIK) SMITHKLINE BEECHAM CORP.

PA (SMIK) SMITHKLINE BEECHAM PLC.

PI Black Mt, Hodgson JE, Knowles DJC, Nicholas RO,

PI Stodola RK;

DR WPI; 97-503111/46.

DR N-PSDB; X30745.

PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in

PT vaccines, drug screening, etc

PS Claim 6; Page 205; 354pp; English.

CC X30724 to X30946 represent genomic DNA sequences isolated from

CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences

CC encode the novel proteins given in Y1114 to Y1137. The proteins,

CC isolated from Streptococcus pneumoniae, can be used in vaccines against

CC streptococcal infections and in assays for identifying compounds that

CC inhibit or activate the activity of the proteins. The antagonists can

CC be used to treat an individual having need to inhibit a bacterial

CC protein. Vectors expressing the proteins can be used to induce a

CC protective immune response in mammals.

CC Sequence 151 AA;

Query Match 42.4%; Score 61; DB 1; Length 151;
 Best Local Similarity 33.3%; Pred. No. 8.99e+01;
 Matches 5; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Db 99 LEEILGIASYPNWP 113

QY 190 MEDVYGPDTYPSDW 204

RESULT	7	
ID	W32621	standard; Protein: 229 AA.
AC	W32621;	
DE	28-JAN-1998	(first entry)
DT	Bacillus smithii	nitrile hydratase subunit beta.
DE	Bacillus smithii	nitrile hydratase subunit alpha; nitrile hydratase subunit beta;
KW	Nitrile hydratase	subunit alpha; nitrile hydratase subunit beta;
KW	acrylonitrile; acrylamide;	biological catalysis; amide;
KW	thermally stable protein.	
OS	Bacillus smithii.	
Key	Location/Qualifiers	
FT	1. .40	
Region	/label= N-terminal_sequence	
ET		
FT		
PN	J09248188-A.	
PD	22-SEP-1997.	
PF	18-MAR-1996; 060732.	
PR	18-MAR-1996; JP-060732.	
PA	(SUMO) SUMITOMO CHEM CO LTD.	

```

Db  · 263 KAMQTVYLREISSQW 278
      |||: || : :|:|
QY  188 KAMEDYGPDTYPSDW 203

```

11 DEWEET HW, EDWEL N, FELT

ET DAYS NO, TOUTIER C, WILLIAMS A,

ET DAYS NO, TOUTHER C, WILLGALISH A;

PI Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM, Shi Y, Young P,
 PI Yu GL;
 DR WPI: 99-080881/07.
 DR N-PSDB: X04380.
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 PS Claim 11; Page 304; 380pp; English.
 CC This sequence represents a secreted human protein encoded by the gene
 CC clone detailed in the descriptor line. The gene can be used to generate
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc
 CC portion (e.g. X04302) for increasing the stability of the fused protein
 CC as compared to the human protein only.
 CC The invention relates to 86 novel genes and their fragments (nucleic acid
 CC sequences: X04311-X04410; amino acid sequences W78126-W78225) which
 CC are useful for preventing, treating or ameliorating medical conditions
 CC e.g. by protein or gene therapy. Also, pathological conditions can be
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the 86 polynucleotides, based on
 CC which tissues they are most highly expressed in (see X04311 for described
 CC uses).
 SQ Sequence 219 AA;

Query Match 41.7%; Score 60; DB 1; Length 219;
 Best Local Similarity 41.2%; Pred. No. 1.11e+02;
 Matches 7; Conservative 4; Mismatches 5; Indels 1; Gaps 1;
 Db 15 AMEQIRYLHEEPESWS 31
 |||::|::|::|::|
 QY 189 AMEDV-YGDPYPSDWS 204

RESULT 12
 ID R07115 standard; protein; 1006 AA.
 AC R07115;
 DT 24-JAN-1991 (first entry)
 DE Gene product of yeast plasmid pVK11.
 KW Beta-galactosidase; lacA+; pVK11.
 OS Aspergillus niger.
 PN W09010703-A.
 PD 20-SEP-1990.
 PF 13-MAR-1990; G00373.
 PR 13-MAR-1989; GB-005674.
 PA (UNLO) IMPERIAL COLLEGE SC.
 PI Hartley BS, Ramakrishnan S, Kumar V;
 DR WPI: 90-305026/40.
 DR N-PSDB; Q06055.
 PT DNA construct contg. beta-galactosidase gene and yeast promoter -
 PT and transformed yeast cells able to ferment lactose in whey
 PS Disclosure; Fig 2; 35pp; English.
 CC Protein is encoded by a Beta-galactosidase lacA+ gene from
 CC Aspergillus niger and the ADL1 yeast promoter in plasmid vector
 CC pVK11. the plasmid is used to transform a yeast expression system to
 CC produce Bg extracellularly or into the periplasmic space, useful in
 CC fermenting lactose to produce EtOH.
 SQ Sequence 1006 AA;

Query Match 41.7%; Score 60; DB 1; Length 1006;
 Best Local Similarity 53.8%; Pred. No. 1.11e+02;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 Db 249 KGAVDIYGHDSYP 261
 |::|::|::|
 QY 188 KAMEDVYGPDTYP 200

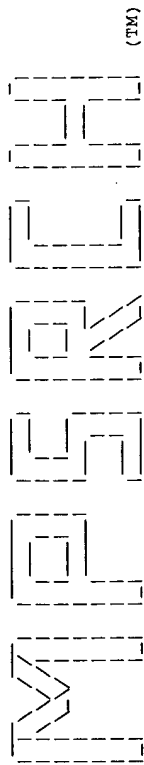
RESULT 13
 ID W17518 standard; Protein; 1007 AA.
 AC W17518;
 DT 21-JUL-1997 (first entry)
 DE Aspergillus niger beta-galactosidase.
 KW Beta-galactosidase; lactose maldigestion; transgenic animal.

OS Aspergillus niger strain A733.
 FH Key Location/Qualifiers
 FT peptide 1..19
 FT cleavage_site 39..40
 FT /label= Sig_peptide
 FT /label= KEX2
 FT /note= "KEX2 cleavage site"
 FT protein 40..1007
 FT /label= Mat_protein
 FT /note= "mature beta-galactosidase"
 FT peptide 820..831
 FT /note= "Characteristic peptide of A. niger A733
 FT beta-galactosidase (Claim 1)"
 FT peptide 903..965
 FT /note= "Characteristic peptide of A. niger A733
 FT beta-galactosidase (Claim 1)"
 PN W09716555-A1.
 PD 09-MAY-1997. CA0723.
 PF 01-NOV-1996; US-551459.
 PR 01-NOV-1995; NEXIA BIOTECHNOLOGIES INC.
 PA (NEXI-) NEXIA BIOTECHNOLOGIES INC.
 PI Delaquis A., Huang Y, Karatzas CN, Lazaris-Karatzas A;
 DR WPI: 97-272123/24.
 DR N-PSDB: T60614-15.
 PT Aspergillus niger beta-galactosidase - used to treat lactose
 PT maldigestion
 PS Claim 3; Page 40-43; 72pp; English.
 CC A beta-galactosidase (W17518) of Aspergillus niger A733 differs
 CC substantially in amino acid sequence from the beta-galactosidase of
 CC A. niger VTT-D-80144. 2 Peptide regions (see also W1800 and
 CC W17517) can be used to characterise the A733 enzyme. The beta-
 CC galactosidase can be expressed in eukaryotic (pref. non-human
 CC mammalian, yeast or fungus, esp. Aspergillus) or prokaryotic (pref.
 CC E. coli) hosts using vectors incorporating isolated cDNA (T60614)
 CC or genomic DNA (T60615) sequences. The beta-galactosidase is used
 CC to treat lactose maldigestion, either by oral administration or
 CC pretreatment of milk. Expression in the mammary glands of a
 CC transgenic animal allows prodn. of low-lactose milk. Non-
 CC glycosylated recombinant beta-galactosidase can be used to produce
 CC antibodies useful in research and quality control.
 SQ Sequence 1007 AA;

Query Match 41.7%; Score 60; DB 1; Length 1007;
 Best Local Similarity 53.8%; Pred. No. 1.11e+02;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 Db 249 KGAVDIYGHDSYP 261
 |::|::|::|
 QY 188 KAMEDVYGPDTYP 200

RESULT 14
 ID R07564 standard; protein; 159 AA.
 AC R07564;
 DT 07-FEB-1991 (first entry)
 DE Polypeptide encoded sal F 3R fragment of Sali F fragment of
 DE the vaccina virus.
 KW Vaccine.
 OS Vaccina virus.
 PN W09012101-A.
 PD 18-OCT-1990.
 PF 03-APR-1990; G00493.
 PR 03-APR-1989; GB-007468.
 PA (LYNX-) LYNXVALE LTD.
 PI Smith GL;
 DR WPI: 90-394851/44.
 DR N-PSDB; Q06358.
 PT Vaccinia virus vectors, genes and expression products - for use
 PT in vaccines.
 PS Disclosure; Fig 5; 78pp; English.
 CC Part or all of the vaccina virus fragment may be removed from
 CC the genome to give a recombinant virus which may have enhanced
 CC safety and immunogenicity over current strains.

Search completed: Wed May 10 13:32:11 2000
Job time : 7 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 13:32:29 2000; MasPar time 44.99 Seconds
Tabular output not generated. 5.472 Million cell updates/sec

Title: >US-09-376-430-2
Description: (188-206) from US09376430A.pep (18 of 25)
Perfect Score: 144
Sequence: 1 KAMEDYGVGDTVPDSEV 19

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 19.962; Variance 70.496; scale 0.283

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	64	44.4	846	1	US-08-451-Sequence 4, Applicatio	2.44e+01
2	64	44.4	846	1	US-08-149-Sequence 4, Applicatio	2.44e+01
3	64	44.4	846	1	US-08-451-Sequence 3, Applicatio	2.44e+01
4	64	44.4	846	1	US-08-149-Sequence 3, Applicatio	2.44e+01
5	64	44.4	873	1	US-08-393-Sequence 2, Applicatio	2.44e+01
6	61	42.4	514	2	US-08-865-Sequence 2, Applicatio	4.63e+01
7	60	41.7	1007	2	US-08-551-Sequence 4, Applicatio	5.72e+01
8	58	40.3	925	2	US-08-504-Sequence 1, Applicatio	8.70e+01
9	58	40.3	925	2	US-08-392-Sequence 1, Applicatio	8.70e+01
10	58	40.3	925	3	PCT-US94-1Sequence 1, Applicatio	8.70e+01
11	57	39.6	911	1	US-08-596-Sequence 2, Applicatio	1.07e+02
12	56	38.9	307	1	US-08-164-Sequence 11, Applicati	1.32e+02
13	56	38.9	307	2	US-08-456-Sequence 11, Applicati	1.32e+02
14	56	38.9	522	1	US-08-164-Sequence 10, Applicati	1.32e+02
15	56	38.9	522	2	US-08-456-Sequence 10, Applicati	1.32e+02
16	56	38.9	536	1	US-08-456-Sequence 12, Applicati	1.32e+02
17	56	38.9	536	2	US-08-164-Sequence 12, Applicati	1.32e+02
18	56	38.9	798	3	PCT-US94-0Sequence 2, Applicatio	1.32e+02
19	56	38.9	798	1	US-08-200-Sequence 2, Applicatio	1.32e+02
20	55	38.2	137	2	US-07-857-Sequence 110, Applicat	1.62e+02
21	55	38.2	944	2	US-08-867-Sequence 24, Applicati	1.62e+02
22	55	38.2	944	2	US-08-867-Sequence 23, Applicati	1.62e+02
23	54	37.5	250	2	US-09-134-Sequence 5, Applicatio	1.99e+02

24	54	37.5	250	2	US-08-861-Sequence 5, Applicatio	1.99e+02
25	54	37.5	629	1	US-07-695-Sequence 2, Applicatio	1.99e+02
26	54	37.5	629	1	US-08-250-Sequence 33, Applicati	1.99e+02
27	54	37.5	630	2	US-08-956-Sequence 3, Applicatio	1.99e+02
28	54	37.5	630	2	US-08-797-Sequence 3, Applicatio	1.99e+02
29	54	37.5	2431	1	US-07-920-Sequence 2, Applicatio	1.99e+02
30	53	36.8	11	3	PCT-US96-0Sequence 9, Applicatio	2.43e+02
31	53	36.8	16	1	US-07-888-Sequence 1, Applicatio	2.43e+02
32	53	36.8	680	1	US-08-211-Sequence 2, Applicatio	2.43e+02
33	53	36.8	1075	2	US-08-993-Sequence 19, Applicati	2.43e+02
34	53	36.8	2500	2	US-08-801-Sequence 2, Applicatio	2.43e+02
35	53	36.8	2512	2	US-08-801-Sequence 9, Applicatio	2.43e+02
36	53	36.8	2517	2	US-08-801-Sequence 5, Applicatio	2.43e+02
37	52	36.1	337	1	US-08-442-Sequence 18, Applicati	2.97e+02
38	52	36.1	555	2	US-08-793-Sequence 34, Applicati	2.97e+02
39	52	36.1	557	2	US-08-793-Sequence 33, Applicati	2.97e+02
40	52	36.1	731	1	US-08-731-Sequence 2, Applicatio	2.97e+02
41	52	36.1	844	1	US-07-731-Sequence 6, Applicatio	2.97e+02
42	52	36.1	844	2	US-08-541-Sequence 6, Applicatio	2.97e+02
43	52	36.1	2186	2	US-08-822-Sequence 2, Applicatio	2.97e+02
44	52	36.1	3672	2	US-08-822-Sequence 12, Applicati	2.97e+02
45	52	36.1	3801	2	US-08-822-Sequence 10, Applicati	2.97e+02

ALIGNMENTS

RESULT 1
ID US-08-451-883-4 STANDARD; PRT; 846 AA.
XX
AC xxxxxx
DT
XX
DE
XX
Sequence 4, Application US/08451883
Sequence 4, Application US/08451883
Patent No. 5798209
GENERAL INFORMATION:
APPLICANT: Lawrence C.B. Chan
TITLE OF INVENTION: HUMAN AND MOUSE VERY LOW DENSITY
TITLE OF INVENTION: LIPOPROTEIN RECEPTORS AND METHODS FOR
TITLE OF INVENTION: USE OF SUCH RECEPTORS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: LYON & LYON
STREET: 633 West Fifth Street, Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: IBM MS-DOS (Version 6.22)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/451,883
FILING DATE: May 26, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below:
APPLICATION NUMBER: 08/149,103
FILING DATE: No. 5798209ember 8, 1993
ATTORNEY/AGENT INFORMATION:
NAME: A.Knight, Matthew W.
REGISTRATION NUMBER: 36,846
REFERENCE/DOCKET NUMBER: 212/268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 4:

```
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 846 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
SQ SEQUENCE 846 AA; 93456 MW; 3603279 CN;

Query Match 44.4%; Score 64; DB 1; Length 846;
Best Local Similarity 37.5%; Pred. No. 2.44e+01;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 480 KMDNVNPAAIADW 495
QY 188 KAMEDVIGDTPSDW 203

RESULT 2
ID US-08-149-103-4 STANDARD; PRT; 846 AA.
XX
AC xxxxxx
DE
DT
XX
XX
XX
Sequence 4, Application US/08149103
CC
CC Sequence 4, Application US/08149103
CC Patent No. 5750367
CC GENERAL INFORMATION:
CC APPLICANT: Lawrence C. B. Chan
CC TITLE OF INVENTION: HUMAN AND MOUSE VERY LOW
CC TITLE OF INVENTION: DENSITY LIPOPROTEIN RECEPTORS
CC TITLE OF INVENTION: AND METHODS FOR USE OF SUCH
CC TITLE OF INVENTION: RECEPTORS
CC NUMBER OF SEQUENCES: 4
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LYON & LYON
CC STREET: 611 West Sixth Street
CC CITY: Los Angeles
CC STATE: California
CC COUNTRY: U.S.A.
CC ZIP: 90017
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
CC SOFTWARE: WordPerfect (Version 5.1)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/149,103
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA: including application
CC PRIOR APPLICATION DATA: described below:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Warburg, Richard J.
CC REGISTRATION NUMBER: 32,327
CC REFERENCE/DOCKET NUMBER: 204/052
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (213) 489-1600
CC TELEFAX: (213) 955-0440
CC TELEX: 67-3510
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 846 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
SQ SEQUENCE 846 AA; 93456 MW; 3603279 CN;

Query Match 44.4%; Score 64; DB 1; Length 846;
Best Local Similarity 37.5%; Pred. No. 2.44e+01;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 846 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
SQ SEQUENCE 846 AA; 93456 MW; 3603279 CN;

Query Match 44.4%; Score 64; DB 1; Length 846;
Best Local Similarity 37.5%; Pred. No. 2.44e+01;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 480 KMDNVNPAAIADW 495
QY 188 KAMEDVIGDTPSDW 203

RESULT 4
ID US-08-149-103-3 STANDARD; PRT; 846 AA.
XX
```

```
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 480 KMDNVNPAAIADW 495
QY 188 KAMEDVIGDTPSDW 203

RESULT 3
ID US-08-451-883-3 STANDARD; PRT; 846 AA.
XX
AC xxxxxx
DE
DT
XX
XX
XX
Sequence 3, Application US/08451883
CC
CC Sequence 3, Application US/08451883
CC Patent No. 5798209
CC GENERAL INFORMATION:
CC APPLICANT: Lawrence C.B. Chan
CC TITLE OF INVENTION: HUMAN AND MOUSE VERY LOW DENSITY
CC TITLE OF INVENTION: LIPOPROTEIN RECEPTORS AND METHODS FOR
CC TITLE OF INVENTION: USE OF SUCH RECEPTORS
CC NUMBER OF SEQUENCES: 4
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LYON & LYON
CC STREET: 633 West Fifth Street, Suite 4700
CC CITY: Los Angeles
CC STATE: California
CC COUNTRY: U.S.A.
CC ZIP: 90071-2066
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: IBM MS-DOS (Version 6.22)
CC SOFTWARE: WordPerfect (Version 5.1)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/451,883
CC FILING DATE: May 26, 1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA: including application
CC PRIOR APPLICATION DATA: described below:
CC APPLICATION NUMBER: 08/149,103
CC FILING DATE: No. 5798209ember 8, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Knight, Matthew W.
CC REGISTRATION NUMBER: 36,846
CC REFERENCE/DOCKET NUMBER: 212/268
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (213) 489-1600
CC TELEFAX: (213) 955-0440
CC TELEX: 67-3510
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 846 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
SQ SEQUENCE 846 AA; 93225 MW; 3613847 CN;

Query Match 44.4%; Score 64; DB 1; Length 846;
Best Local Similarity 37.5%; Pred. No. 2.44e+01;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 480 KMDNVNPAAIADW 495
QY 188 KAMEDVIGDTPSDW 203

RESULT 4
ID US-08-149-103-3 STANDARD; PRT; 846 AA.
XX
```

xxxxxx

AC
XX
DT
XX
DE
XX
Sequence 3, Application US/08149103
CC
XX
Sequence 3, Application US/08149103
CC
Patent No. 5750367
CC
GENERAL INFORMATION:
CC
APPLICANT: Lawrence C. B. Chan
CC
TITLE OF INVENTION: HUMAN AND MOUSE VERY LOW
CC
DENSITY LIPOPROTEIN RECEPTORS
CC
TITLE OF INVENTION: AND METHODS FOR USE OF SUCH
CC
RECEPTORS
CC
NUMBER OF SEQUENCES: 4
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: LYON & LYON
CC
STREET: 611 West Sixth Street
CC
CITY: Los Angeles
CC
STATE: California
CC
COUNTRY: U.S.A.
CC
ZIP: 90017
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
CC
COMPUTER: IBM PC compatible
CC
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
CC
SOFTWARE: WordPerfect (Version 5.1)
CC
CURRENT APPLICATION DATA:
CC
APPLICATION NUMBER: US/08/149,103
CC
FILING DATE:
CC
CLASSIFICATION: 435
CC
PRIOR APPLICATION DATA: including application
CC
PRIOR APPLICATION DATA: described below:
CC
APPLICATION NUMBER:
CC
FILING DATE:
CC
ATTORNEY/AGENT INFORMATION:
CC
NAME: Warburg, Richard J.
CC
REGISTRATION NUMBER: 32,327
CC
REFERENCE/DOCKET NUMBER: 204/052
CC
TELECOMMUNICATION INFORMATION:
CC
TELEPHONE: (213) 489-1600
CC
TELEFAX: (213) 955-0440
CC
TELEX: 67-3510
CC
INFORMATION FOR SEQ ID NO: 3:
CC
SEQUENCE CHARACTERISTICS:
CC
TYPE: amino acid
CC
STRANDEDNESS: single
CC
TOPOLOGY: linear
CC
SEQUENCE 846 AA; 93225 MW; 3613847 CN;
SQ

Query Match 44.4%; Score 64; DB 1; Length 846;
Best Local Similarity 37.5%; Pred. No. 2.44e+01;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 480 KMDNVNPAIAVDW 495
QY 188 KAMEDVIGDTPSDW 203
| : : : | : : : |

RESULT 5
ID US-08-393-734-2 STANDARD; PRT; 873 AA.

xxxxxx

AC
XX
DT
XX
DE
XX
Sequence 2, Application US/08393734
CC
XX
Sequence 2, Application US/08393734
CC
Patent No. 5652224
CC
GENERAL INFORMATION:
CC

CC
XX
DT
XX
DE
XX
Sequence 3, Application US/08149103
CC
XX
Sequence 3, Application US/08149103
CC
Patent No. 5750367
CC
GENERAL INFORMATION:
CC
APPLICANT: Lawrence C. B. Chan
CC
TITLE OF INVENTION: HUMAN AND MOUSE VERY LOW
CC
DENSITY LIPOPROTEIN RECEPTORS
CC
TITLE OF INVENTION: AND METHODS FOR USE OF SUCH
CC
RECEPTORS
CC
NUMBER OF SEQUENCES: 4
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: LYON & LYON
CC
STREET: 611 West Sixth Street
CC
CITY: Los Angeles
CC
STATE: California
CC
COUNTRY: U.S.A.
CC
ZIP: 90017
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
CC
COMPUTER: IBM PC compatible
CC
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
CC
SOFTWARE: WordPerfect (Version 5.1)
CC
CURRENT APPLICATION DATA:
CC
APPLICATION NUMBER: US/08/149,103
CC
FILING DATE:
CC
CLASSIFICATION: 435
CC
PRIOR APPLICATION DATA: including application
CC
PRIOR APPLICATION DATA: described below:
CC
APPLICATION NUMBER:
CC
FILING DATE:
CC
ATTORNEY/AGENT INFORMATION:
CC
NAME: Warburg, Richard J.
CC
REGISTRATION NUMBER: 32,327
CC
REFERENCE/DOCKET NUMBER: 204/052
CC
TELECOMMUNICATION INFORMATION:
CC
TELEPHONE: (213) 489-1600
CC
TELEFAX: (213) 955-0440
CC
TELEX: 67-3510
CC
INFORMATION FOR SEQ ID NO: 3:
CC
SEQUENCE CHARACTERISTICS:
CC
TYPE: amino acid
CC
STRANDEDNESS: single
CC
TOPOLOGY: linear
CC
SEQUENCE 846 AA; 93225 MW; 3613847 CN;
SQ

Query Match 44.4%; Score 64; DB 1; Length 846;
Best Local Similarity 37.5%; Pred. No. 2.44e+01;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 507 KMDNVNPAIAVDW 522
QY 188 KAMEDVIGDTPSDW 203
| : : : | : : : |

RESULT 6
ID US-08-865-311-2 STANDARD; PRT; 514 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 2, Application US/08865311
CC
XX
Sequence 2, Application US/08865311
CC
Patent No. 5919664
CC
GENERAL INFORMATION:
CC
APPLICANT: Pearson, Stewart C.
CC
TITLE OF INVENTION: No. 5919664el Compounds
CC
NUMBER OF SEQUENCES: 2
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: Dechert, Price & Rhoads
CC
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CC
CITY: Philadelphia
CC
STATE: PA
CC
COUNTRY: USA
CC
ZIP: 19103-2793
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Diskette
CC
COMPUTER: IBM Compatible
CC
OPERATING SYSTEM: Windows 95
CC

APPLICANT: Wilson, James M.
CC
APPLICANT: Kozarsky, Karen F.
CC
APPLICANT: Strauss, Jerome F.
CC
TITLE OF INVENTION: Methods and Compositions for Gene
CC
TITLE OF INVENTION: Therapy for the Treatment of Defects in Lipoprotein
CC
TITLE OF INVENTION: Metabolism
CC
NUMBER OF SEQUENCES: 8
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: Howson and Howson
CC
STREET: Spring House Corporate Cntr., PO Box 457
CC
CITY: Spring House
CC
STATE: Pennsylvania
CC
COUNTRY: USA
CC
ZIP: 19477
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Floppy disk
CC
COMPUTER: IBM PC compatible
CC
OPERATING SYSTEM: PC-DOS/MS-DOS
CC
SOFTWARE: Patentin Release #1.0, Version #1.30
CC
CURRENT APPLICATION DATA:
CC
APPLICATION NUMBER: US/08/393,734
CC
FILING DATE:
CC
CLASSIFICATION: 424
CC
ATTORNEY/AGENT INFORMATION:
CC
NAME: Bak, Mary E.
CC
REGISTRATION NUMBER: 31,215
CC
REFERENCE/DOCKET NUMBER: UPNH1254USA
CC
TELECOMMUNICATION INFORMATION:
CC
TELEPHONE: 215-540-9200
CC
TELEFAX: 215-540-5818
CC
INFORMATION FOR SEQ ID NO: 2:
CC
SEQUENCE CHARACTERISTICS:
CC
LENGTH: 873 amino acids
CC
TYPE: amino acid
CC
TOPOLOGY: linear
CC
MOLECULE TYPE: protein
CC
SEQUENCE 873 AA; 96121 MW; 3833800 CN;
SQ

FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 08632/005001
TELECOMMUNICATION INFORMATION:

CC SEQUENCE CHARACTERISTICS:

Matches 8; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

AC xxxxxxxx

DT XX
DE XX
XX
Sequence 2, Application US/08596985
Sequence 2, Application US/08596985
Patent No. 5736374
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Hucul, John A.
APPLICANT: Ward, Michael
TITLE OF INVENTION: Increased production of
TITLE OF INVENTION: Beta-galactosidase in *Aspergillus oryzae*
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International, Inc
STREET: 180 Kimball Way
CITY: South San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596,985
FILING DATE: 05-FEB-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/267,631
FILING DATE: 29-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Horn, Margaret A
REGISTRATION NUMBER: 33,401
REFERENCE/DOCKET NUMBER: GC250
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 742-7536
TELEFAX: (415) 742-7217
INFORMATION FOR SEQ ID NO: 2:
LENGTH: 911 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 911 AA; 99238 MW; 4569084 CN;
Query Match 39.6%; Score 57; DB 1; Length 911;
Best Local Similarity 66.7%; Pred. No. 1.07e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 253 DIYGHDSYP 261
Qy 192 DVIYGPDTYP 200
RESULT 12
ID US-08-164-614A-11 STANDARD; PRT; 307 AA.
XX
AC xxxxxx
XX
DT
DE
XX
Sequence 11, Application US/08164614A
Sequence 11, Application US/08164614A
Patent No. 5789237
GENERAL INFORMATION:
APPLICANT: Renauld, Jean-Christophe
APPLICANT: Druetz, Catherine
APPLICANT: Van Snick, Jacques
TITLE OF INVENTION: Nucleic Acid Sequences Coding For

CC
CC
CC
TITLE OF INVENTION: Or Complementary To Nucleic Acid Sequences Coding For
TITLE OF INVENTION: Interleukin 9 Receptor
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/164,614A
FILING DATE: 8-DECEMBER-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/847,347
FILING DATE: 09-MARCH-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5789237man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5264.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acid residues
TYPE: amino acid
TOPOLOGY: linear
SEQUENCE 307 AA; 34256 MW; 521435 CN;
Query Match 38.9%; Score 56; DB 1; Length 307;
Best Local Similarity 40.0%; Pred. No. 1.32e+02;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
Db 233 DDVVEERYTCQWSE 247
Qy 191 EDVYGPDTYPSDWSE 205
RESULT 13
ID US-08-456-489B-11 STANDARD; PRT; 307 AA.
XX
AC xxxxxx
XX
DT
DE
XX
Sequence 11, Application US/08456489B
Sequence 11, Application US/08456489B
Patent No. 5962269
GENERAL INFORMATION:
APPLICANT: Renauld, Jean-Christophe; Druetz, Catherine; Van Snick,
APPLICANT: Jacques
TITLE OF INVENTION: Nucleic Acid Sequences Coding For Or
TITLE OF INVENTION: Complementary To Nucleic Acid Sequences Coding For Inte
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski L.L.P.
STREET: 666 Fifth Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS

SQ SEQUENCE 522 AA; 57319 MW; 1432274 CN;

Query Match 38.9%; Score 56; DB 2; Length 522;
Best Local Similarity 40.0%; Pred. NO. 1.32e+02;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 233 DDVVEERYTGOWSE 247

:||:|:|:|

QY 191 EDVIGPDTPSDWSE 205

Search completed: Wed May 10 13:33:24 2000
Job time : 55 secs.

W P S R E L I (TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 13:31:39 2000; Maspar time 4.56 Seconds
Tabular output not generated. 196.630 Million cell updates/sec

Title: >US-09-376-430-2
Description: (188-206) from US09376430A.pap (18 of 25)
Perfect Score: 144
Sequence: 1 KAMEDVYGPDPYPSDWSEV 19

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 29.009; Variance 45.890; scale 0.632

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	69	47.9	298	2	S69523 hypothetical protein	7.51e+01
2	68	47.2	102	2	D69288 hypothetical protein	1.11e+00
3	67	46.5	676	2	D72749 probable cleavage and	1.63e+00
4	67	46.5	4436	2	E71086 hypothetical protein	1.63e+00
5	66	45.8	255	2	E72226 conserved hypothetical	2.38e+00
6	66	45.1	60	2	JC1389 homeotic protein Eghb	3.47e+00
7	64	44.4	873	1	A49729 VLDL receptor precurs	5.04e+00
8	64	44.4	873	1	I48952 VLDL receptor precurs	5.04e+00
9	64	44.4	873	1	QRRBVD VLDL receptor precurs	5.04e+00
10	64	44.4	1036	2	H64245 hypothetical protein	5.04e+00
11	63	43.8	157	2	S25479 homeotic protein Hox7	7.29e+00
12	63	43.8	672	2	E72391 beta-D-galactosidase	7.29e+00
13	62	43.1	198	2	QJ0864 hypothetical 21.9K pr	1.05e+01
14	62	43.1	205	2	S53318 6-phosphofructokinase	1.05e+01
15	62	43.1	212	2	R69317 6-phosphofructokinase	1.05e+01
16	62	43.1	344	2	R69895 phosphoglycerate dehy	1.05e+01
17	62	43.1	424	2	S11676 spore coat protein SP	1.05e+01
18	62	43.1	780	1	KIRBF 6-phosphofructokinase	1.05e+01
19	62	43.1	780	1	KIHUFM 6-phosphofructokinase	1.05e+01
20	62	43.1	780	2	S71429 phosphofructokinase	1.05e+01
21	61	42.4	332	2	C64164 thiamine-binding peri	1.51e+01
22	61	42.4	449	2	H72458 probable tryptophan s	1.51e+01
23	61	42.4	488	2	E37202 xylan 1,4-beta-xylosi	1.51e+01

24	61	42.4	609	2	S36569	E1 protein - human pa	1.51e+01
25	61	42.4	932	2	C70191	penicillin-binding pr	1.51e+01
26	61	42.4	2493	2	S26372	nonstructural polypro	1.51e+01
27	61	42.4	2493	2	S72349	nonstructural polypro	1.51e+01
28	60	41.7	110	2	T03913	hypothetical protein	2.16e+01
29	60	41.7	137	2	JQ2323	virion protein - Afri	2.16e+01
30	60	41.7	459	2	G71036	probable tryptophan s	2.16e+01
31	60	41.7	619	1	S48729	glucokinase regulator	2.16e+01
32	60	41.7	1082	2	T15269	hypothetical protein	2.16e+01
33	59	41.0	159	2	JQ1777	Salp2R protein precur	3.07e+01
34	59	41.0	168	2	F42521	A40R protein - vaccin	3.07e+01
35	59	41.0	367	2	S28604	St12p protein - Arabi	3.07e+01
36	59	41.0	381	2	F64354	hypothetical protein	3.07e+01
37	59	41.0	480	3	T08415	putative mitochondria	3.07e+01
38	59	41.0	504	2	T05325	hypothetical protein	3.07e+01
39	59	41.0	578	2	A70877	probable acyl-coA synt	3.07e+01
40	59	41.0	605	2	S36592	E1 protein - human pa	3.07e+01
41	59	41.0	688	2	F64111	DNA-directed DNA pol	3.07e+01
42	59	41.0	841	2	T01011	hypothetical protein	3.07e+01
43	59	41.0	2514	1	MNWN2	nonstructural polypro	3.07e+01
44	58	40.3	869	1	JC4858	VLDL receptor precurs	4.35e+01
45	58	40.3	925	1	A39216	plasma cell membrane	4.35e+01

ALIGNMENTS

RESULT 1
ENTRY #type complete
TITLE hypothetical protein 17 - phage HPI
ORGANISM #formal_name phage HPI
DATE 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
ACCESSIONS S69523
REFERENCE Espósito, D.; Fitzmaurice, W.P.; Benjamin, R.C.; Goodman, S.D.; Waldman, A.S.; Scocca, J.J.
#authors Nucleic Acids Res. (1996) 24:2360-2368
#journal The complete nucleotide sequence of bacteriophage HPI DNA.
#title
#cross-references EMBL:U24159; MID:g1046235; PID:g1046244
#accession S69523
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-298 #label ESP
#cross-references EMBL:U24159; MID:g1046235; PID:g1046244
#note the nucleotide sequence was submitted to the EMBL Data Library, April 1995

SUMMARY #length 298 #molecular-weight 33702 #checksum 3998

Query Match 47.9%; Score 69; DB 2; Length 298;
Best Local Similarity 35.3%; Pred. No. 7.51e+01;
Matches 6; Conservative 5; Mismatches 5; Indels 1; Gaps 1;
Db 33 MAETYPDEHYTANLWPE 49
QY 190 MEDVYGPDPYPSDWSE 205

RESULT 2
ENTRY #type complete
TITLE hypothetical protein AF0308 - Archaeoglobus fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 03-Dec-1997 #text_change 05-Jun-1998
ACCESSIONS D69288
REFERENCE 'A69250
#authors Klenk, H.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeill, L.K.; Badger, J.H.;

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Pyrococcus horikoshii* fulgidus.
#cross-references MUID:98049343
#accession D69288
#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA
#residues 1-102 #label KLE
#cross-references GB:AE001083; GB:AE000782; NID:g2689406; PID:g2650328; TIGR:AF0308

SUMMARY #length 102 #molecular-weight 12265 #checksum 4565

Query Match 47.2%; Score 68; DB 2; Length 102;
Best Local Similarity 58.3%; Pred. No. 1.11e+00;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 29 DIFIPDEVPKDW 40
QY 192 DVYGPDTYPSDW 203

RESULT 3
ENTRY C72749 #type complete
TITLE probable cleavage and polyadenylation factor subunit APE0522

ORGANISM - Aeropyrum pernix (strain K1)
DATE 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change

ACCESSIONS C72749
REFERENCE A72450

#authors Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Anka, I.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.

#journal DNA Res. (1999) 6:83-101
#title Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyrum pernix* K1.

#cross-references MUID:99310339
#accession C72749
#status preliminary

#residues 1-676 #label KAW
#cross-references DBJ:AP000059; NID:g5103911; PIDN:BAA79487.1; PID:d1043273; PID:g5104171

#experimental_source strain K1
GENETICS #gene APE0522
SUMMARY #length 676 #molecular-weight 76444 #checksum 9165

Query Match 46.5%; Score 67; DB 2; Length 676;
Best Local Similarity 72.7%; Pred. No. 1.63e+00;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 233 YGPDSPYVYWS 243
QY 194 YGPDYPSDWS 204

RESULT 4
ENTRY E71086 #type complete
TITLE hypothetical protein PH0954 - *Pyrococcus horikoshii*

ORGANISM #formal_name *Pyrococcus horikoshii*
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change

ACCESSIONS E71086
REFERENCE A71000

#authors Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.

#journal DNA Res. (1998) 5:55-76
#title Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.

#cross-references MUID:98344137
#accession E71086
#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA
#residues 1-4436 #label KAW
#cross-references GB:AP000004; NID:g3236131; PID:d1030994; PID:g3257368

#experimental_source strain OT3
#note this accession replaces an interim accession for a sequence replaced by GenBank

GENETICS #gene PH0954
SUMMARY #length 4436 #molecular-weight 497021 #checksum 7775

Query Match 46.5%; Score 67; DB 2; Length 4436;
Best Local Similarity 41.7%; Pred. No. 1.63e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 1113 IYSTNAYPRNWS 1124
QY 193 YVGPDYPSDWS 204

RESULT 5
ENTRY E72226 #type complete
TITLE conserved hypothetical protein - *Thermotoga maritima* (strain MSB8)

ORGANISM #formal_name *Thermotoga maritima*
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change

ACCESSIONS E72226
REFERENCE A72200

#authors Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek, J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

#journal Nature (1999) 399:323-329
#title Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*.

#cross-references MUID:99287316
#accession E72226
#status preliminary

#molecule_type DNA
#residues 1-255 #label ARN
#cross-references GB:AE001808; GB:AE000512; NID:g4982233; PID:g4982254; TIGR:TM1679

#experimental_source strain MSB8
GENETICS #gene TM1679
SUMMARY #length 255 #molecular-weight 28679 #checksum 9494

Query Match 45.8%; Score 66; DB 2; Length 255;
Best Local Similarity 36.8%; Pred. No. 2.38e+00;
Matches 7; Conservative 8; Mismatches 3; Indels 1; Gaps 1;

Db 80 ALDQYSGRYAGADWNEV 98

```

|:::|:|:|:|:|:|
QY 189 AMEDVYGPDPYP-SDWSEV 206

RESULT 6
ENTRY JCI1389 #type fragment
TITLE homeotic protein Egrbx4 - tapeworm (Echinococcus granulosus)
ORGANISM #formal_name Echinococcus granulosus
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Oct-1997
ACCESSIONS JCI1389
REFERENCE JCI1386
#authors Oliver, G.; Vispo, M.; Mailhos, A.; Martinez, C.; Sosa-Pineda, B.; Fiehlitz, W.; Ehrlich, R.
#journal Gene (1992) 121:337-342
#title Homeoboxes in flatworms.
#cross-references MUID:93077050
#accession JCI1389
#molecule_type DNA
#residues 1-60 #label OLI
#cross-references EMBL:X66820
CLASSIFICATION #superfamily unassigned homeobox proteins; homeobox homology
KEYWORDS DNA binding; homeobox; nucleus; transcription regulation
FEATURE
2-58 #domain homeobox homology #label HOX
SUMMARY #length 60 #checksum 7663
Query Match 45.1%; Score 65; DB 2; Length 60;
Best Local Similarity 58.3%; Pred. No. 3.47e+00;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 15 AMEEVFGVNRYP 26
|:::|:|:|:|:|
QY 189 AMEDVYGPDPYP 200

RESULT 7
ENTRY A49729 #type complete
TITLE VLDL receptor precursor, long splice form - human
ALTERNATE_NAMES very low density lipoprotein receptor; vitellogenin receptor
CONTAINS VLDL receptor short splice form
ORGANISM #formal_name Homo sapiens #common_name man
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 22-Jun-1999
ACCESSIONS A49729; B49729; A54309; I54373; I59603
REFERENCE A49729
#authors Sakai, J.; Hoshino, A.; Takahashi, S.; Miura, Y.; Ishii, H.; Suzuki, H.; Kawarabayashi, Y.; Yamamoto, T.
#journal J. Biol. Chem. (1994) 269:2173-2182
#title Structure, chromosome location, and expression of the human very low density lipoprotein receptor gene.
#cross-references MUID:94124575
#accession A49729
#status preliminary
#molecule_type mRNA
#residues 1-873 #label SAK
#cross-references GB:D16493; NID:g391733; PIDN:BAA03945.1; PID:d1004461; PID:g391734
#accession B49729
#status preliminary
#molecule_type mRNA
#residues 1-750,779-873 #label SA2
#cross-references GB:D16494
REFERENCE A54309
#authors Oka, K.; Tzung, K.W.; Sullivan, M.; Lindsay, E.; Baldini, A.; Chan, L.
#journal Genomics (1994) 20:298-300
#title Human very-low-density lipoprotein receptor complementary DNA and deduced amino acid sequence and localization of its gene (VLDLR) to chromosome band 9p24 by fluorescence in situ hybridization.
#cross-references MUID:94292216
#accession A54309
#status preliminary
#molecule_type mRNA
#residues 1-12, 'V', 14-765, 'S', 767-873 #label OKA
#cross-references GB:L22431; NID:g437386; PIDN:AAA61344.1; PID:g437387
REFERENCE I54373
#authors Webb, J.C.; Patel, D.D.; Jones, M.D.; Knight, B.L.; Soutar, A.K.
#journal Hum. Mol. Genet. (1994) 3:531-537
#title Characterization and tissue-specific expression of the human very low density lipoprotein (VLDL) receptor' mRNA.
#cross-references MUID:94348496
#accession I54373
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type mRNA
#residues 1-873 #label RES
#cross-references GB:S73849; NID:g688370; PIDN:AA831735.1; PID:g688371
REFERENCE I59603
#authors Gafvels, M.E.; Caird, M.; Britt, D.; Jackson, C.L.; Patterson, D.; Strauss, J.F.
#journal Somat. Cell Mol. Genet. (1993) 19:557-569
#title Cloning of a cDNA encoding a putative human very low density lipoprotein/apolipoprotein E receptor and assignment of the gene to chromosome 9pter-p23.
#cross-references MUID:94174378
#accession I59603
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type mRNA
#residues 1-8, 'V', 10-423, 'A', 425-677, 'H', 679-873 #label RE2
#cross-references GB:L20470; NID:g409425; PIDN:AAA53684.1; PID:g409426
GENETICS
#gene GDB:VLDLR
#cross-references GDB:219649; OMIM:192977
#map_position 9p24-9p24
CLASSIFICATION #superfamily LDL receptor; EGF homology; LDL receptor ligand-binding repeat homology; LDL receptor YWTD-containing repeat homology; alternative splicing; duplication; fatty acid metabolism; glycoprotein; receptor; transmembrane protein; VLDL
KEYWORDS
FEATURE
1-27 #domain signal sequence #status predicted #label SIG\
28-873 #product VLDL receptor, long splice form #status predicted #label MAT\
28-797 #domain extracellular #status predicted #label EXT\
28-750,779-873 #product VLDL receptor, short splice form #status predicted #label MAT2\
33-67 #domain LDL receptor ligand-binding repeat homology #label LDL1\
72-108 #domain LDL receptor ligand-binding repeat homology #label LDL2\
113-149 #domain LDL receptor ligand-binding repeat homology #label LDL3\
154-188 #domain LDL receptor ligand-binding repeat homology #label LDL4\
193-229 #domain LDL receptor ligand-binding repeat homology #label LDL5\
239-273 #domain LDL receptor ligand-binding repeat homology #label LDL6\
278-312 #domain LDL receptor ligand-binding repeat homology #label LDL7\
318-355 #domain LDL receptor ligand-binding repeat homology #label LDL8\
360-394 #domain EGF homology #label EG1\
400-434 #domain EGF homology #label EG2\
441-480 #domain LDL receptor YWTD-containing repeat homology #label YW1\
481-524 #domain LDL receptor YWTD-containing repeat homology #label YW2\
525-567 #domain LDL receptor YWTD-containing repeat homology #label YW3\
568-611 #domain LDL receptor YWTD-containing repeat homology #label YW4\
612-654 #domain LDL receptor YWTD-containing repeat homology

```

```

655-697      #label YW5\
#domain LDL receptor YWTD-containing repeat homology
#label YW6\
706-749      #domain EGF homology #label EG3\
751-790      #region clustered O-linked oligosaccharides\
798-819      #domain transmembrane #status predicted #label TM\
834-838      #region coated-pit mediated internalization signal\
151,765,781  #binding_site carbohydrate (Asn) (covalent) #status
              predicted\
360-371,367-380,
382-394,400-410,
406-419,421-434,
706-719,715-734,
736-749      #disulfide_bonds #status predicted
              #length 873 #molecular_weight 96097 #checksum 4691
SUMMARY
Query Match 44.4%; Score 64; DB 1; Length 873;
Best Local Similarity 37.5%; Pred. No. 5.04e+00;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
Db 507 KMTDNYNPAIAVDW 522
QY 188 KAMEDVGPDTYPSDW 203
      |:::| |:::|

RESULT 8
ENTRY I48952 #type complete
TITLE VLDL receptor precursor, long splice form - mouse
ALTERNATE_NAMES very low density lipoprotein receptor; vitellogenin receptor
CONTAINS VLDL receptor short splice form
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
ACCESSIONS I48952; S48217
REFERENCE I48952
#authors Gafvels, M.E.; Paavola, L.G.; Boyd, C.O.; Nolan, P.M.;
Wittmaack, F.; Chavira, A.; Lazar, M.A.; Bucan, M.; Angelin,
B.O.; Strauss, J.F.
#journal Endocrinology (1994) 135:387-394
#title Cloning of a complementary deoxyribonucleic acid encoding the
murine homolog of the very low density
lipoprotein/apolipoprotein-E receptor: expression pattern
and assignment of the gene to mouse chromosome 19.
#cross-references MUID:94283285
#accession I48952
##status Preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-873 #label RES
##cross-references EMBL:U06670; NID:g619646; PIDN:AAA59384.1;
PID:g619647
REFERENCE S48217
#authors Oka, K.; Ishimura-Oka, K.; Chu, M.; Sullivan, M.; Krushkal,
J.; Li, W.H.; Chan, L.
#journal Eur. J. Biochem. (1994) 224:975-982
#title Mouse very-low-density-lipoprotein receptor (VLDLR) cDNA
cloning, tissue-specific expression and evolutionary
relationship with the low-density-lipoprotein receptor.
#cross-references MUID:95010090
#accession S48217
##status Preliminary
##molecule_type mRNA
##residues 1-296,'C',298-873 #label OKA
##cross-references GB:I33417; NID:g609532
CLASSIFICATION #superfamily LDL receptor; EGF homology; LDL receptor
ligand-binding repeat homology; LDL receptor
YWTD-containing repeat homology
alternative splicing; duplication; fatty acid metabolism;
glycoprotein; receptor; transmembrane protein; VLDL
KEYWORDS #domain signal sequence #status predicted #label SIG\
#product VLDL receptor, long splice form #status
1-27 Predicted #label MAT\
28-873 #domain extracellular #status predicted #label EXT\
28-797

```

```

28-750,779-873
33-67      #product VLDL receptor, short splice form #status
              predicted #label MAT2\
#domain LDL receptor ligand-binding repeat homology
#label LDL1\
72-108      #domain LDL receptor ligand-binding repeat homology
#label LDL2\
113-149      #domain LDL receptor ligand-binding repeat homology
#label LDL3\
154-188      #domain LDL receptor ligand-binding repeat homology
#label LDL4\
193-229      #domain LDL receptor ligand-binding repeat homology
#label LDL5\
239-273      #domain LDL receptor ligand-binding repeat homology
#label LDL6\
278-312      #domain LDL receptor ligand-binding repeat homology
#label LDL7\
318-355      #domain LDL receptor ligand-binding repeat homology
#label LDL8\
360-394      #domain EGF homology #label EG1\
400-434      #domain EGF homology #label EG2\
441-480      #domain LDL receptor YWTD-containing repeat homology
#label YW1\
481-524      #domain LDL receptor YWTD-containing repeat homology
#label YW2\
525-567      #domain LDL receptor YWTD-containing repeat homology
#label YW3\
568-611      #domain LDL receptor YWTD-containing repeat homology
#label YW4\
612-654      #domain LDL receptor YWTD-containing repeat homology
#label YW5\
655-697      #domain LDL receptor YWTD-containing repeat homology
#label YW6\
706-749      #domain EGF homology #label EG3\
751-790      #region clustered O-linked oligosaccharides\
798-819      #domain transmembrane #status predicted #label TM\
834-838      #region coated-pit mediated internalization signal\
151,765,781  #binding_site carbohydrate (Asn) (covalent) #status
              predicted\
360-371,367-380,
382-394,400-410,
406-419,421-434,
706-719,715-734,
736-749      #disulfide_bonds #status predicted
              #length 873 #molecular_weight 96356 #checksum 3530
SUMMARY
Query Match 44.4%; Score 64; DB 1; Length 873;
Best Local Similarity 37.5%; Pred. No. 5.04e+00;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
Db 507 KMTDNYNPAIAVDW 522
QY 188 KAMEDVGPDTYPSDW 203
      |:::| |:::|

RESULT 9
ENTRY QRBVD #type complete
TITLE VLDL receptor precursor - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
              rabbit
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
ACCESSIONS 16-Jul-1999
REFERENCE A46286
#authors Takahashi, S.; Kawarabayashi, Y.; Nakai, T.; Sakai, J.;
#journal Proc. Natl. Acad. Sci. U.S.A. (1992) 89:9252-9256
#title Rabbit very low density lipoprotein receptor: a low density
lipoprotein receptor-like protein with distinct ligand
specificity.
#cross-references MUID:93028442
#accession A46286
##molecule_type mRNA
##residues 1-873 #label TAK

```

```

##cross-references DBJ:D11100; NID:g217752; PID:BAA01874.1;
PID:d1002351; PID:g217753
##note sequence extracted from NCBI backbone (NCBIN:115853;
NCBIP:115854)
COMMENT This protein is abundant in heart, muscle, and adipose tissue.
CLASSIFICATION #superfamily LDL receptor; EGF homology; LDL receptor
ligand-binding repeat homology; LDL receptor
YWTD-containing repeat homology
cholesterol; coated pits; duplication; endocytosis;
glycoprotein; lipid transport; receptor; transmembrane
protein; VLDL

FEATURE
1-27 #domain signal sequence #status predicted #label SIG
28-873 #product VLDL receptor #status predicted #label MAT\
28-797 #domain extracellular #status predicted #label EXT\
33-67 #domain LDL receptor ligand-binding repeat homology
#label LDL1\
72-108 #domain LDL receptor ligand-binding repeat homology
#label LDL2\
113-149 #domain LDL receptor ligand-binding repeat homology
#label LDL3\
154-198 #domain LDL receptor ligand-binding repeat homology
#label LDL4\
193-229 #domain LDL receptor ligand-binding repeat homology
#label LDL5\
239-273 #domain LDL receptor ligand-binding repeat homology
#label LDL6\
278-312 #domain LDL receptor ligand-binding repeat homology
#label LDL7\
318-355 #domain LDL receptor ligand-binding repeat homology
#label LDL8\
360-394 #domain EGF homology #label EG1\
400-434 #domain EGF homology #label EG2\
441-480 #domain LDL receptor YWTD-containing repeat homology
#label YW1\
481-524 #domain LDL receptor YWTD-containing repeat homology
#label YW2\
525-567 #domain LDL receptor YWTD-containing repeat homology
#label YW3\
568-611 #domain LDL receptor YWTD-containing repeat homology
#label YW4\
612-654 #domain LDL receptor YWTD-containing repeat homology
#label YW5\
655-697 #domain LDL receptor YWTD-containing repeat homology
#label YW6\
706-749 #domain EGF homology #label EG3\
751-790 #region clustered O-linked oligosaccharides\
798-819 #domain transmembrane #status predicted #label TMW\
820-873 #domain intracellular #status predicted #label INT\
834-838 #region coated-pit mediated internalization signal\
151,765,781 #binding_site carbohydrate (Asn) (covalent) #status
predicted\
360-371,367-380,
382-394,400-410,
406-419,421-434,
706-719,715-734,
736-749 #disulfide_bonds #status predicted
#length 873 #molecular_weight 96280 #checksum 5713
SUMMARY
Query Match 44.4%; Score 64; DB 1; Length 873;
Best Local Similarity 37.5%; Pred. No. 5.04e+00;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps

Db 507 KMIDNVYNPAIAVDW 522
|:::| |:::|
QY 188 KAMEDVYGPDTYPSDW 203

RESULT 10
ENTRY H64245 #type complete
TITLE hypothetical protein MG414 - Mycoplasma genitalium (SGC3)
ORGANISM #formal_name Mycoplasma genitalium
DATE 10-Nov-1995 #sequence revision 10-Nov-1995 #text_change

```

```

26-Aug-1999
E72391
A72200
#accessions
#reference
#authors
Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson,
R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson,
W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek,
J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton,
M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.;
Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
#journal
#title
Nature (1999) 398:323-329
Evidence for lateral gene transfer between Archaea and
Bacteria from genome sequence of Thermotoga maritima.
#cross-references MUID:99287316
#accession E72391
#status preliminary
#molecule_type DNA
#residues 1-672 #label ARN
#cross-references GB:AE000713; GB:AE000512; NID:g4980809; PID:g4980811;
TIGR:TM0310
#experimental_source strain MSB8
GENETICS
#gene TM0310
#superfamily Bacillus beta-galactosidase
CLASSIFICATION
#length 672 #molecular_weight 79112 #checksum 6645
SUMMARY
Query Match 43.8%; Score 63; DB 2; Length 672;
Best Local Similarity 58.3%; Pred. No. 7.29e+00;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Db 11 YGGDYYPEQWDE 22
||| ||| |||
QY 194 YGPDTPDSE 205

RESULT 13
ENTRY J00864 #type complete
TITLE hypothetical 21.9k protein - Escherichia coli retron Ec67
ALTERNATE_NAMES ci protein homolog
ORGANISM Escherichia coli retron Ec67
DATE 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change
10-Sep-1997
J00864
#accessions J00851
#reference Hsu, M.Y.; Inouye, M.; Inouye, S.
#authors Proc. Natl. Acad. Sci. U.S.A. (1990) 87:9454-9458
#journal Retron for the 67-base multicopy single-stranded DNA from
#title Escherichia coli: a potential transposable element encoding
both reverse transcriptase and Dam methylase functions.
#cross-references MUID:91067724
#accession J00864
#molecule_type DNA
#residues 1-198 #label HSU
#cross-references GB:M55249; NID:g145143; PID:g145144
#experimental_source E. coli strain Cl-1
GENETICS
#note insertion site is equivalent to 19 min of E. coli K12 genetic
map
SUMMARY
#length 198 #molecular_weight 21898 #checksum 7751
Query Match 43.1%; Score 62; DB 2; Length 198;
Best Local Similarity 40.0%; Pred. No. 1.05e+01;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
Db 47 TLANYLRDTPPADW 61
::: ||| |||
QY 189 AMEDVYGPDPDSE 203

RESULT 14
ENTRY S53318 #type fragment
TITLE 6-phosphofructokinase (EC 2.7.1.11) M - rat (fragment)
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat

```

```

15-Jul-1995 #sequence_revision 10-May-1996 #text_change
07-May-1999
S53318
#accessions S53317
#reference Nakajima, H.; Noguchi, T.; Hamaguchi, T.; Tomita, K.;
#authors Hanafusa, T.; Kono, N.; Tanaka, T.; Kuwajima, M.;
Matsuzawa, Y.
#journal Biochem. J. (1994) 303:449-453
#title Expression of mouse phosphofructokinase-M gene alternative
transcripts: evidence for the conserved two-promoter
system.
#cross-references MUID:95071243
#accession S53318
#status translation not shown
#molecule_type mRNA
#residues 1-205 #label NAK
#cross-references GB:D21869; NID:g567260; PID:d1005419; PID:g1255681
#experimental_source strain SD
GENETICS
#gene PFK-M
#superfamily human 6-phosphofructokinase;
CLASSIFICATION 6-phosphofructokinase 1 homology
#keywords ATP; phosphotransferase
FEATURE
11-205
SUMMARY
#domain 6-phosphofructokinase 1 homology (fragment)
#length 205 #checksum 9408
Query Match 43.1%; Score 62; DB 2; Length 205;
Best Local Similarity 50.0%; Pred. No. 1.05e+01;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 117 GADTFRSEWSDL 128
||||| ||| |||
QY 195 GPDTPDSEV 206

RESULT 15
ENTRY S53317 #type fragment
TITLE 6-phosphofructokinase (EC 2.7.1.11) M - mouse (fragment)
ALTERNATE_NAMES #formal_name Mus musculus #common_name house mouse
ORGANISM 15-Jul-1995 #sequence_revision 10-May-1996 #text_change
07-May-1999
S53317
#accessions S53317
#reference Nakajima, H.; Noguchi, T.; Hamaguchi, T.; Tomita, K.;
#authors Hanafusa, T.; Kono, N.; Tanaka, T.; Kuwajima, M.;
Matsuzawa, Y.
#journal Biochem. J. (1994) 303:449-453
#title Expression of mouse phosphofructokinase-M gene alternative
transcripts: evidence for the conserved two-promoter
system.
#cross-references MUID:95071243
#accession S53317
#status translation not shown
#molecule_type mRNA
#residues 1-212 #label NAK
#cross-references GB:D21865; NID:g567254; PID:d1005418; PID:g1255637
GENETICS
#gene PFK-M
#superfamily human 6-phosphofructokinase;
CLASSIFICATION 6-phosphofructokinase 1 homology
#keywords ATP; phosphotransferase
FEATURE
18-212
SUMMARY
#domain 6-phosphofructokinase 1 homology (fragment)
#length 212 #checksum 9975
Query Match 43.1%; Score 62; DB 2; Length 212;
Best Local Similarity 50.0%; Pred. No. 1.05e+01;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 124 GADTFRSEWSDL 135

```

Thu May 11 06:50:05 2000

US-09-376-430-2-18.rpr

Page 7

QY 195 GPDTPSDWSEV 206
|:|:|: |:|:|:

Search completed: Wed May 10 13:31:48 2000
Job time : 9 secs.

W P E R L E H

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 13:25:20 2000; Master time 90.60 seconds
Tabular output not generated. 6.387 Million cell updates/sec

Title: >US-09-376-430-2
Description: (188-206) from US09376430A.pap (18 of 25)
Perfect Score: 144
Sequence: 1 KAMEDVYGPDTYPSDMSV 19

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 29.708; Variance 43.122; scale 0.689

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	69	47.9	298	1	VPO_BPHP1	2.85e+01
2	65	45.1	60	1	HBX4_ECHGR	1.46e+00
3	64	44.4	873	1	LDVR_MOUSE	2.17e+00
4	64	44.4	873	1	LDVR_RAT	2.17e+00
5	64	44.4	873	1	LDVR_HUMAN	2.17e+00
6	64	44.4	873	1	LDVR_MOUSE	2.17e+00
7	64	44.4	957	1	SECA_MYCSE	2.17e+00
8	64	44.4	1036	1	Y414_MYCGE	2.17e+00
9	62	43.1	183	1	CRVP_TRIMU	4.74e+00
10	62	43.1	198	1	IR71_ECOLI	4.74e+00
11	62	43.1	211	1	K6PF_MOUSE	4.74e+00
12	62	43.1	424	1	SP60_DICDI	4.74e+00
13	62	43.1	779	1	K6PF_RAT	4.74e+00
14	62	43.1	779	1	K6PF_HUMAN	4.74e+00
15	62	43.1	779	1	K6PF_MOUSE	4.74e+00
16	62	43.1	781	1	K6PF_CANFA	4.74e+00
17	61	42.4	332	1	TPBA_HAEIN	6.97e+00
18	61	42.4	435	1	ORCS_HUMAN	6.97e+00
19	61	42.4	488	1	YXNB_CALSA	6.97e+00
20	61	42.4	609	1	VEL_HPV49	6.97e+00
21	60	41.7	223	1	PRHL_MOUSE	1.02e+01
22	60	41.7	226	1	PRHL_HUMAN	1.02e+01
23	60	41.7	619	1	GCKR_XENLA	1.02e+01

24	60	41.7	1006	1	BGAL_ASPNG	1.02e+01
25	59	41.0	168	1	VA40_VACCC	1.48e+01
26	59	41.0	168	1	VA40_VACCV	1.48e+01
27	59	41.0	381	1	Y438_METJA	1.48e+01
28	59	41.0	605	1	VEL_HPV09	1.48e+01
29	59	41.0	688	1	DP3X_HAEIN	1.48e+01
30	59	41.0	2514	1	POLN_ONNVG	1.48e+01
31	58	40.3	119	1	YRF4_SHIFL	2.15e+01
32	58	40.3	131	1	YJGH_ECOLI	2.15e+01
33	58	40.3	206	1	HS27_RAT	2.15e+01
34	58	40.3	246	1	YBSC_GLOHI	2.15e+01
35	58	40.3	302	1	RIR2_EBV	2.15e+01
36	58	40.3	327	1	TPBA_ECOLI	2.15e+01
37	58	40.3	405	1	Y872_METJA	2.15e+01
38	58	40.3	428	1	Y958_HUMAN	2.15e+01
39	58	40.3	612	1	PEX5_YEAST	2.15e+01
40	58	40.3	735	1	DHDM_HYPSX	2.15e+01
41	58	40.3	873	1	PCI_HUMAN	2.15e+01
42	57	39.6	270	1	YNK8_YEAST	3.09e+01
43	57	39.6	430	1	YHT2_YEAST	3.09e+01
44	57	39.6	440	1	SCAL_DROME	3.09e+01
45	57	39.6	461	1	D15K_MOUSE	3.09e+01

ALIGNMENTS

RESULT 1
ID VPO_BPHP1 STANDARD; PRT: 298 AA.
AC P51719;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PROBABLE CAPSID SCAFFOLDING PROTEIN (ORF17).
OS Bacteriophage HP1.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae.
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=HPIC1;
RX MEDLINE; 96279738.
RA Esposito D., Fitzmaurice W.P., Benjamin R.C., Goodman S.D., Waldman A.S., Scocca J.J.
RT "The complete nucleotide sequence of bacteriophage HP1 DNA."
RL Nucleic Acids Res. 24:2360-2368(1996).
CC -!- SIMILARITY: STRONG, TO PHAGE P2 PROTEIN O.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL: U24159; AAB09202.1;
CC Capsid assembly.
DR EMBL; U24159; AAB09202.1;
KW Capsid assembly.
SQ SEQUENCE 298 AA; 33702 MW; 6D2841468D700C1F CRC64;

Query Match 47.9%; Score 69; DB 1; Length 298;
Best Local Similarity 35.3%; Pred. No. 2.85e+01;
Matches 6; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db	33	MAETYPDEHYTANLWPE	49
QY	190	MEDVYGPDTYPSD-WSE	205
RESULT 2			
ID HBX4_ECHGR STANDARD; PRT: 60 AA.			
AC P55813;			
DT 01-NOV-1997 (Rel. 35, Created)			
DT 01-NOV-1997 (Rel. 35, Last sequence update)			
DT 15-JUL-1999 (Rel. 38, Last annotation update)			
DE HOMEBOX PROTEIN EGBHX4 (FRAGMENT).			


```

FT DISULFID 113 127 BY SIMILARITY.
FT DISULFID 124 140 BY SIMILARITY.
FT DISULFID 134 149 BY SIMILARITY.
FT DISULFID 154 166 BY SIMILARITY.
FT DISULFID 161 179 BY SIMILARITY.
FT DISULFID 173 188 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 200 218 BY SIMILARITY.
FT DISULFID 212 229 BY SIMILARITY.
FT DISULFID 232 251 BY SIMILARITY.
FT DISULFID 246 264 BY SIMILARITY.
FT DISULFID 258 273 BY SIMILARITY.
FT DISULFID 278 290 BY SIMILARITY.
FT DISULFID 285 303 BY SIMILARITY.
FT DISULFID 297 312 BY SIMILARITY.
FT DISULFID 318 331 BY SIMILARITY.
FT DISULFID 326 344 BY SIMILARITY.
FT DISULFID 338 355 BY SIMILARITY.
FT DISULFID 360 371 BY SIMILARITY.
FT DISULFID 367 380 BY SIMILARITY.
FT DISULFID 382 394 BY SIMILARITY.
FT DISULFID 400 410 BY SIMILARITY.
FT DISULFID 406 419 BY SIMILARITY.
FT DISULFID 421 434 BY SIMILARITY.
FT DISULFID 706 719 BY SIMILARITY.
FT DISULFID 715 734 BY SIMILARITY.
FT DISULFID 736 749 BY SIMILARITY.
FT CARBOHYD 151 151 POTENTIAL.
FT CARBOHYD 765 765 POTENTIAL.
FT CARBOHYD 781 781 POTENTIAL.
FT CONFLICT 161 161 C -> G (IN REF. 1).
FT CONFLICT 262 262 P -> L (IN REF. 3).
FT CONFLICT 297 297 C -> S (IN REF. 2).
SQ SEQUENCE 873 AA; 96372 MW; 08F09F93825195CB CRC64;

Query Match 44.4%; Score 64; DB 1; Length 873;
Best Local Similarity 37.5%; Pred. No. 2.17e+00;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 507 KNIDNVNPAIAVDW 522
Qy 188 KAMEDVYGDTPYSDW 203

RESULT 4 STANDARD; PRT; 873 AA.
ID LDVR_RAT
AC P98166;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE VERY LOW-DENSITY LIPOPROTEIN RECEPTOR PRECURSOR (VLDL RECEPTOR).
GN VLDLR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 95014480.
RA Jokinen E.V., Landschulz K.T., Wyne K.L., Ho Y.K., Frykman P.K.,
RA Hobbs H.H.;
RT "Regulation of the very low density lipoprotein receptor by thyroid
RT hormone in rat skeletal muscle."
RL J. Biol. Chem. 269:26411-26418(1994).
CC -!- FUNCTION: BINDS VLDL AND TRANSPORTS IT INTO CELLS BY ENDOCYTOSIS.
CC IN ORDER TO BE INTERNALIZED, THE RECEPTOR-LIGAND COMPLEXES MUST
CC FIRST CLUSTER INTO CLATHRIN-COATED PITS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: ABUNDANT IN SOLEUS, GASTROCNEMIUS, HEART
CC MUSCLE, PLACENTA, BRAIN, LUNG AND WHITE FAT. LESS IN AORTA, OVARY,
CC KIDNEY, SPLEEN, ADRENAL GLAND, AND THYMUS.
CC -!- SIMILARITY: CONTAINS 8 LDL-RECEPTOR CLASS A DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 LDL-RECEPTOR CLASS B DOMAINS.

```

-!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; L35767; AAA42341.1; -
 RSP; P01130; IAUJ.
 PRINTS; PR0261; LDLRECEPTOR.
 PROSITE; PS0010; ASX_HYDROXYL; 2.
 PROSITE; PS00022; EGF_1; FALSE_NEG.
 PROSITE; PS01186; EGF_2; 3.
 PROSITE; PS01187; EGF_CA; 1.
 PROSITE; PS01209; LDLRA_1; 8.
 PROSITE; PS00068; LDLRA_2; 8.
 PFAM; PF00057; ldl_recept_a; 8.
 PFAM; PF00058; ldl_recept_b; 5.
 Glycoprotein; VLDL; Cholesterol metabolism; Lipid transport;
 Endocytosis; Coated pits; Transmembrane; Receptor; Signal;
 EGF-like domain; Repeat.
 SIGNAL 1 27
 CHAIN 28 873
 DOMAIN 28 797
 TRANSMEM 798 819
 DOMAIN 820 873
 DOMAIN 31 69
 DOMAIN 70 110
 DOMAIN 111 151
 DOMAIN 152 190
 DOMAIN 191 231
 DOMAIN 237 275
 DOMAIN 276 314
 DOMAIN 316 355
 DOMAIN 356 395
 DOMAIN 396 435
 REPEAT 439 480
 REPEAT 481 524
 REPEAT 525 567
 REPEAT 568 611
 REPEAT 612 654
 REPEAT 655 696
 DOMAIN 702 750
 DOMAIN 751 790
 SITE 837 837
 DISULFID 33 45
 DISULFID 40 58
 DISULFID 52 67
 DISULFID 72 84
 DISULFID 79 97
 DISULFID 91 108
 DISULFID 113 127
 DISULFID 120 140
 DISULFID 134 149
 DISULFID 154 166
 DISULFID 161 179
 DISULFID 173 188
 DISULFID 193 205
 DISULFID 200 218
 DISULFID 212 229
 DISULFID 239 251
 DISULFID 246 264
 DISULFID 258 273
 DISULFID 278 290
 DISULFID 285 303
 DISULFID 297 312
 DISULFID 318 331
 DISULFID 326 344
 DISULFID 338 355
 DISULFID 360 371

-!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; L35767; AAA42341.1; -
 RSP; P01130; IAUJ.
 PRINTS; PR0261; LDLRECEPTOR.
 PROSITE; PS0010; ASX_HYDROXYL; 2.
 PROSITE; PS00022; EGF_1; FALSE_NEG.
 PROSITE; PS01186; EGF_2; 3.
 PROSITE; PS01187; EGF_CA; 1.
 PROSITE; PS01209; LDLRA_1; 8.
 PROSITE; PS00068; LDLRA_2; 8.
 PFAM; PF00057; ldl_recept_a; 8.
 PFAM; PF00058; ldl_recept_b; 5.
 Glycoprotein; VLDL; Cholesterol metabolism; Lipid transport;
 Endocytosis; Coated pits; Transmembrane; Receptor; Signal;
 EGF-like domain; Repeat.
 SIGNAL 1 27
 CHAIN 28 873
 DOMAIN 28 797
 TRANSMEM 798 819
 DOMAIN 820 873
 DOMAIN 31 69
 DOMAIN 70 110
 DOMAIN 111 151
 DOMAIN 152 190
 DOMAIN 191 231
 DOMAIN 237 275
 DOMAIN 276 314
 DOMAIN 316 355
 DOMAIN 356 395
 DOMAIN 396 435
 REPEAT 439 480
 REPEAT 481 524
 REPEAT 525 567
 REPEAT 568 611
 REPEAT 612 654
 REPEAT 655 696
 DOMAIN 702 750
 DOMAIN 751 790
 SITE 837 837
 DISULFID 33 45
 DISULFID 40 58
 DISULFID 52 67
 DISULFID 72 84
 DISULFID 79 97
 DISULFID 91 108
 DISULFID 113 127
 DISULFID 120 140
 DISULFID 134 149
 DISULFID 154 166
 DISULFID 161 179
 DISULFID 173 188
 DISULFID 193 205
 DISULFID 200 218
 DISULFID 212 229
 DISULFID 239 251
 DISULFID 246 264
 DISULFID 258 273
 DISULFID 278 290
 DISULFID 285 303
 DISULFID 297 312
 DISULFID 318 331
 DISULFID 326 344
 DISULFID 338 355
 DISULFID 360 371

-!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; L35767; AAA42341.1; -
 RSP; P01130; IAUJ.
 PRINTS; PR0261; LDLRECEPTOR.
 PROSITE; PS0010; ASX_HYDROXYL; 2.
 PROSITE; PS00022; EGF_1; FALSE_NEG.
 PROSITE; PS01186; EGF_2; 3.
 PROSITE; PS01187; EGF_CA; 1.
 PROSITE; PS01209; LDLRA_1; 8.
 PROSITE; PS00068; LDLRA_2; 8.
 PFAM; PF00057; ldl_recept_a; 8.
 PFAM; PF00058; ldl_recept_b; 5.
 Glycoprotein; VLDL; Cholesterol metabolism; Lipid transport;
 Endocytosis; Coated pits; Transmembrane; Receptor; Signal;
 EGF-like domain; Repeat.
 SIGNAL 1 27
 CHAIN 28 873
 DOMAIN 28 797
 TRANSMEM 798 819
 DOMAIN 820 873
 DOMAIN 31 69
 DOMAIN 70 110
 DOMAIN 111 151
 DOMAIN 152 190
 DOMAIN 191 231
 DOMAIN 237 275
 DOMAIN 276 314
 DOMAIN 316 355
 DOMAIN 356 395
 DOMAIN 396 435
 REPEAT 439 480
 REPEAT 481 524
 REPEAT 525 567
 REPEAT 568 611
 REPEAT 612 654
 REPEAT 655 696
 DOMAIN 702 750
 DOMAIN 751 790
 SITE 837 837
 DISULFID 33 45
 DISULFID 40 58
 DISULFID 52 67
 DISULFID 72 84
 DISULFID 79 97
 DISULFID 91 108
 DISULFID 113 127
 DISULFID 120 140
 DISULFID 134 149
 DISULFID 154 166
 DISULFID 161 179
 DISULFID 173 188
 DISULFID 193 205
 DISULFID 200 218
 DISULFID 212 229
 DISULFID 239 251
 DISULFID 246 264
 DISULFID 258 273
 DISULFID 278 290
 DISULFID 285 303
 DISULFID 297 312
 DISULFID 318 331
 DISULFID 326 344
 DISULFID 338 355
 DISULFID 360 371

-!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; L35767; AAA42341.1; -
 RSP; P01130; IAUJ.
 PRINTS; PR0261; LDLRECEPTOR.
 PROSITE; PS0010; ASX_HYDROXYL; 2.
 PROSITE; PS00022; EGF_1; FALSE_NEG.
 PROSITE; PS01186; EGF_2; 3.
 PROSITE; PS01187; EGF_CA; 1.
 PROSITE; PS01209; LDLRA_1; 8.
 PROSITE; PS00068; LDLRA_2; 8.
 PFAM; PF00057; ldl_recept_a; 8.
 PFAM; PF00058; ldl_recept_b; 5.
 Glycoprotein; VLDL; Cholesterol metabolism; Lipid transport;
 Endocytosis; Coated pits; Transmembrane; Receptor; Signal;
 EGF-like domain; Repeat.
 SIGNAL 1 27
 CHAIN 28 873
 DOMAIN 28 797
 TRANSMEM 798 819
 DOMAIN 820 873
 DOMAIN 31 69
 DOMAIN 70 110
 DOMAIN 111 151
 DOMAIN 152 190
 DOMAIN 191 231
 DOMAIN 237 275
 DOMAIN 276 314
 DOMAIN 316 355
 DOMAIN 356 395
 DOMAIN 396 435
 REPEAT 439 480
 REPEAT 481 524
 REPEAT 525 567
 REPEAT 568 611
 REPEAT 612 654
 REPEAT 655 696
 DOMAIN 702 750
 DOMAIN 751 790
 SITE 837 837
 DISULFID 33 45
 DISULFID 40 58
 DISULFID 52 67
 DISULFID 72 84
 DISULFID 79 97
 DISULFID 91 108
 DISULFID 113 127
 DISULFID 120 140
 DISULFID 134 149
 DISULFID 154 166
 DISULFID 161 179
 DISULFID 173 188
 DISULFID 193 205
 DISULFID 200 218
 DISULFID 212 229
 DISULFID 239 251
 DISULFID 246 264
 DISULFID 258 273
 DISULFID 278 290
 DISULFID 285 303
 DISULFID 297 312
 DISULFID 318 331
 DISULFID 326 344
 DISULFID 338 355
 DISULFID 360 371

RESULT	6	
ID	LDVR_HUMAN	STANDARD; PRT: 873 AA.

FT DISULFID 367 380 BY SIMILARITY.
 FT DISULFID 382 394 BY SIMILARITY.
 FT DISULFID 400 410 BY SIMILARITY.
 FT DISULFID 406 419 BY SIMILARITY.
 FT DISULFID 421 434 BY SIMILARITY.
 FT DISULFID 706 719 BY SIMILARITY.
 FT DISULFID 715 734 BY SIMILARITY.
 FT DISULFID 736 749 BY SIMILARITY.
 FT CARBOHYD 151 151 POTENTIAL.
 FT CARBOHYD 765 765 POTENTIAL.
 FT CARBOHYD 781 781 POTENTIAL.
 FT VARSPLIC 751 779 STARTVYSEKONTNTEISATSGLVPGG -> R (IN SHORT ISOFORM).
 FT CONFLICT 9 9 L -> V (IN REF. 1).
 FT CONFLICT 13 13 L -> V (IN REF. 4).
 FT CONFLICT 424 424 G -> A (IN REF. 1).
 FT CONFLICT 678 678 L -> H (IN REF. 1).
 FT CONFLICT 766 766 T -> S (IN REF. 4).
 SQ SEQUENCE 873 AA; 96098 MW; 8BAC29438A78C2B8 CRC64;
 Query Match 44.4%; Score 64; DB 1; Length 873;
 Best Local Similarity 37.5%; Pred. No. 2.17e+00;
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 Db 507 KMDINVYNPAIAVDW 522
 QY 188 KAMEDVYGPDTYPSDW 203
 RESULT 7
 ID SECA_MYCSM STANDARD; PRT; 957 AA.
 AC P71533;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE PREPROTEIN TRANSLOCASE SECA SUBUNIT.
 GN SECA.
 OS Mycobacterium smegmatis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Brown A.B., Jacobs W.R.; to the EMBL/GenBank/DBJ databases.
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: INVOLVED IN PROTEIN EXPORT. INTERACTS WITH THE SECY/SECE
 CC SUBUNITS. SECA HAS A CENTRAL ROLE IN COUPLING THE HYDROLYSIS OF
 CC ATP TO THE TRANSFER OF PRE-SECRETORY PERIPLASMIC AND OUTER
 CC MEMBRANE PROTEINS ACROSS THE MEMBRANE (BY SIMILARITY).
 CC -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY
 CC (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC SIDE OF MEMBRANE
 CC (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SECA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U66081; AAB06754.1; -;
 CC PRINTS; PRO0906; SECA.
 CC DR PROSITE; PS01312; SECA; 1.
 CC DR PFAM; PF01043; SecA_protein; 1.
 CC KW Protein transport; ATP-binding; Membrane; Translocation; Transport.
 CC FT NP_BIND 98 105 ATP (POTENTIAL).
 CC SQ SEQUENCE: 957 AA; 107001 MW; AD208569A228A32E CRC64;
 Query Match 44.4%; Score 64; DB 1; Length 957;
 Best Local Similarity 31.6%; Pred. No. 2.17e+00;
 Matches 6; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

Matches 6; Conservative 6; Mismatches 7; Indels 0; Gaps 0;
 Db 512 RGLDPVETPEEYAAWHEV 530
 QY 188 KAMEDVYGPDTYPSDWSEV 206
 RESULT 8
 ID Y414_MYCGE STANDARD; PRT; 1036 AA.
 AC P47653; P47654; Q49457;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE HYPOTHETICAL PROTEIN MG414.
 GN MG414.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 33530 / G-37;
 RX MEDLINE; 96046346.
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
 RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RT "The minimal gene complement of Mycoplasma genitalium";
 RL Science 270:397-403(1995).
 RN [2]
 RP SEQUENCE OF 52-146 AND 733-833 FROM N.A.
 RC STRAIN-ATCC 33530 / G-37;
 RX MEDLINE; 94075230.
 RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
 RT "A survey of the Mycoplasma genitalium genome by using random
 RT sequencing";
 RL J. Bacteriol. 175:7918-7930(1993).
 CC -!- SIMILARITY: BELONGS TO THE MG414 / MG415 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U39723; AAC71641.1; -;
 CC EMBL; U01695; AAB01008.1; -;
 CC DR TIGR; MG414; -;
 CC KW Hypothetical protein; Transmembrane.
 CC FT TRANSMEM 4 24 POTENTIAL.
 CC FT TRANSMEM 1004 1024 POTENTIAL.
 CC FT CONFLICT 733 736 LEPT -> SRAS (IN REF. 2).
 CC SQ SEQUENCE 1036 AA; 123179 MW; 12A21F00F686A141 CRC64;
 Query Match 44.4%; Score 64; DB 1; Length 1036;
 Best Local Similarity 31.6%; Pred. No. 2.17e+00;
 Matches 6; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

Db 69 RPNQNNYFAKYSNEWSEL 87
 QY 188 KAMEDVYGPDTYPSDWSEV 206
 RESULT 9
 ID CRVP_TRIMU STANDARD; PRT; 183 AA.
 AC P79845;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE CYSTEINE-RICH VENOM PROTEIN PRECURSOR (CRVP).
OS Trimerus mucroscumatus (Taiwan habu).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Lepidosauria;
OC Squamata; Scieroglossa; Serpentes; Colubroidea; Viperidae; Crotalinae;
OC Trimerus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=VENOM GLAND;
RA Chang T.Y., Mao S., Guo Y.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U59447; AAB48565.1; -
DR PRINTS; P00837; V5TPXLIKE.
DR PROSITE; P001009; SCP_AG5_P1_SC7_1; 1.
DR PROSITE; P001010; SCP_AG5_P1_SC7_2; 1.
DR PFAM; P00188; SCP; 1.
DR Signal.
KW CHAIN
FT SIGNAL
FT CHAIN
FT SIGNAL
SQ SEQUENCE 183 AA; 20378 MW; 50224D8E92A64930 CRC64;

Query Match 43.1%; Score 62; DB 1; Length 183;
Best Local Similarity 31.3%; Pred. No. 4.74e+00;
Matches 5; Conservative 7; Mismatches 3; Indels 1; Gaps 1;

Db 37 ENIYMS-PYPAKWTDI 51
QY 191 EDVGYPTDTPSDWSEV 206
::: ||: ||:

RESULT 10
ID YR7I.ECOLI STANDARD; PRT; 198 AA.
AC P21323;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAY-1991 (Rel. 18, Last annotation update)
DE HYPOTHETICAL 21.9 KD PROTEIN (ORFI) (RETRON EC67).
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CL-1;
RX MEDLINE; 91067724.
RA Hsu M.-Y., Inouye M., Inouye S.;
RT "Retron for the 67-base multicopy single-stranded DNA from
RT Escherichia coli: a potential transposable element encoding both
RT reverse transcriptase and Dam methylase functions.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9454-9458(1990).
CC - SIMILARITY: 30% IDENTITY TO CI OF BACTERIOPHAGE 186.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M55249; JAA23390.1; -
DR PIR; J00864; J00864
KW Transposable element; Hypothetical protein.
SQ SEQUENCE 198 AA; 21898 MW; 14CED3124E28516B CRC64;

Query Match 43.1%; Score 62; DB 1; Length 198;
Best Local Similarity 40.0%; Pred. No. 4.74e+00;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 47 TLNRYLRDTPADW 61
QY 189 AMEDVGYPTDTPSDW 203
::: ||: ||:

RESULT 11
ID K6PF_MOUSE STANDARD; PRT; 211 AA.
AC P47857;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 6-PHOSPHOFRUCTOKINASE, MUSCLE TYPE (EC 2.7.1.11) (PHOSPHOFRUCTOKINASE
DE 1) (PHOSPHOHEXOKINASE) (PHOSPHOFRUCTO-1-KINASE ISOZYME A) (FRAGMENT).
GN PFKM OR PFK-M.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR;
RX MEDLINE; 95071243.
RA Nakajima H., Noguchi T., Hamaguchi T., Tomita K., Hanafusa T.,
RA Kono N., Tanaka T., Kuwajima M., Matsuzawa Y.;
RT "Expression of mouse phosphofructokinase-M gene alternative
RT transcripts: evidence for the conserved two-promoter system";
RL Biochem. J. 303:449-453(1994).
CC - CATALYTIC ACTIVITY: ATP + D-FRUCTOSE 6-PHOSPHATE = ADP +
CC D-FRUCTOSE 1,6-BISPHOSPHATE.
CC - COFACTOR: REQUIRES MAGNESIUM.
CC - ENZYME REGULATION: ALLOSTERIC ENZYME ACTIVATED BY ADP, AMP, OR
CC FRUCTOSE BISPHOSPHATE AND INHIBITED BY ATP OR CITRATE.
CC - PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
CC - SUBUNIT: HOMOTETRAMER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D21865; BAA21012.1; -
DR HSP; P06998; 2PFK.
DR MGD; MGI:97548; PFKM.
DR PROSITE; P000433; PHOSPHOFRUCTOKINASE; PARTIAL.
DR PFAM; P00365; PFK; 1.
DR Kinase; Transferase; Glycolysis; Repeat; Allosteric enzyme;
KW Phosphorylation; Magnesium; Multigene family.
FT INIT_MET 0
FT NON_TER 211 211
SQ SEQUENCE 211 AA; 22639 MW; B7B3EFA6C79A5601 CRC64;

Query Match 43.1%; Score 62; DB 1; Length 211;
Best Local Similarity 50.0%; Pred. No. 4.74e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 123 GADFRSEWSDL 134
QY 195 GPDTPSDWSEV 206
::: ||: ||:

RESULT 12
ID SP60_DICDI STANDARD; PRT; 424 AA.
AC P15270;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)

DE CYSTEINE-RICH VENOM PROTEIN PRECURSOR (CRVP).
OS Trimerus mucroscumatus (Taiwan habu).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Lepidosauria;
OC Squamata; Scieroglossa; Serpentes; Colubroidea; Viperidae; Crotalinae;
OC Trimerus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=VENOM GLAND;
RA Chang T.Y., Mao S., Guo Y.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U59447; AAB48565.1; -
DR PRINTS; P00837; V5TPXLIKE.
DR PROSITE; P001009; SCP_AG5_P1_SC7_1; 1.
DR PROSITE; P001010; SCP_AG5_P1_SC7_2; 1.
DR PFAM; P00188; SCP; 1.
DR Signal.
KW CHAIN
FT SIGNAL
FT CHAIN
FT SIGNAL
SQ SEQUENCE 183 AA; 20378 MW; 50224D8E92A64930 CRC64;

Query Match 43.1%; Score 62; DB 1; Length 183;
Best Local Similarity 31.3%; Pred. No. 4.74e+00;
Matches 5; Conservative 7; Mismatches 3; Indels 1; Gaps 1;

Db 37 ENIYMS-PYPAKWTDI 51
QY 191 EDVGYPTDTPSDWSEV 206
::: ||: ||:

RESULT 10
ID YR7I.ECOLI STANDARD; PRT; 198 AA.
AC P21323;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAY-1991 (Rel. 18, Last annotation update)
DE HYPOTHETICAL 21.9 KD PROTEIN (ORFI) (RETRON EC67).
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CL-1;
RX MEDLINE; 91067724.
RA Hsu M.-Y., Inouye M., Inouye S.;
RT "Retron for the 67-base multicopy single-stranded DNA from
RT Escherichia coli: a potential transposable element encoding both
RT reverse transcriptase and Dam methylase functions.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9454-9458(1990).
CC - SIMILARITY: 30% IDENTITY TO CI OF BACTERIOPHAGE 186.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M55249; JAA23390.1; -
DR PIR; J00864; J00864
KW Transposable element; Hypothetical protein.
SQ SEQUENCE 198 AA; 21898 MW; 14CED3124E28516B CRC64;

RL J. Biol. Chem. 262:4195-4199(1987).
RN [2]
RX SEQUENCE.
RP MEDLINE: 84219739.
RA Poorman R.A., Randolph A., Kemp R.G., Heinrikson R.L.;
RT "Evolution of phosphofructokinase -- gene duplication and creation of
RT new effector sites."
RL Nature 309:467-469(1984).
CC -1- CATALYTIC ACTIVITY: ATP + D-FRUCTOSE 6-PHOSPHATE = ADP +
CC D-FRUCTOSE 1,6-BISPHOSPHATE.
CC -1- COFACTOR: REQUIRES MAGNESIUM.
CC -1- ENZYME REGULATION: ALLOSTERIC ENZYME ACTIVATED BY ADP, AMP, OR
CC FRUCTOSE BISPHOSPHATE AND INHIBITED BY ATP OR CITRATE.
CC -1- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14477; AAA31441.1; JOINED.
DR EMBL; M14456; AAA31441.1; JOINED.
DR EMBL; M14457; AAA31441.1; JOINED.
DR EMBL; M14458; AAA31441.1; JOINED.
DR EMBL; M14459; AAA31441.1; JOINED.
DR EMBL; M14460; AAA31441.1; JOINED.
DR EMBL; M14461; AAA31441.1; JOINED.
DR EMBL; M14462; AAA31441.1; JOINED.
DR EMBL; M14463; AAA31441.1; JOINED.
DR EMBL; M14464; AAA31441.1; JOINED.
DR EMBL; M14465; AAA31441.1; JOINED.
DR EMBL; M14466; AAA31441.1; JOINED.
DR EMBL; M14467; AAA31441.1; JOINED.
DR EMBL; M14468; AAA31441.1; JOINED.
DR EMBL; M14469; AAA31441.1; JOINED.
DR EMBL; M14470; AAA31441.1; JOINED.
DR EMBL; M14471; AAA31441.1; JOINED.
DR EMBL; M14472; AAA31441.1; JOINED.
DR EMBL; M14473; AAA31441.1; JOINED.
DR EMBL; M14474; AAA31441.1; JOINED.
DR EMBL; M14475; AAA31441.1; JOINED.
DR EMBL; M14476; AAA31441.1; JOINED.
DR PIR; A00604; KIRBF.
DR PIR; A26550; A26550.
DR HSSP; P00512; 6PFK.
DR PRINTS; PR00476; PHFRCTKINASE.
DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 2.
DR PFAM; PF00365; PFK; 2.
KW Kinase; Transferase; Glycolysis; Repeat; Allosteric enzyme;
KW Phosphorylation; Magnesium; Multigene family.
FT INIT_MET 0
FT REPEAT 1 401 APPROXIMATE.
FT REPEAT 402 779 APPROXIMATE.
FT MOD_RES 774 774 PHOSPHORYLATION.
FT CONFLICT 268 268 R -> S (IN REF. 2).
FT CONFLICT 479 507 MISSING (IN REF. 2).
FT CONFLICT 558 558 S -> I (IN REF. 2).
FT CONFLICT 565 565 MISSING (IN REF. 2).
SQ SEQUENCE 779 AA; 85072 MW; 96D8E7C7B5D3C8A9 CRC64;

Query Match 43.1%; Score 62; DB 1; Length 779;
Best Local Similarity 50.0%; Pred. No. 4.74e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 123 GADTFRSEWSDL 134
Qy 195 GPDTPSDWSEV 206
I::I::I::I::

RESULT 15
ID K6PF_HUMAN STANDARD; PRT; 779 AA.
AC P08237; Q16814;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 6-PHOSPHOFRUCTOKINASE, MUSCLE TYPE (EC 2.7.1.11) (PHOSPHOFRUCTOKINASE
DE 1) (PHOSPHOHEXOKINASE) (PHOSPHOFRUCTO-1-KINASE ISOZYME A).
GN PFKM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=MUSCLE;
RX MEDLINE: 92009225.
RA Yamasaki T., Nakajima H., Kono N., Hotta K., Yamada K., Imai E.,
RA Kuwajima M., Noguchi T., Tanaka T., Tarui S.;
RT "Structure of the entire human muscle phosphofructokinase-encoding
RT gene: a two-promoter system."
RL Gene 104:277-282(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=MUSCLE;
RX MEDLINE: 89306675.
RA Sharma P.M., Reddy G.R., Vora S., Babior B.M., McLachlan A.;
RT "Cloning and expression of a human muscle phosphofructokinase cDNA."
RL Gene 77:177-183(1989).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=MUSCLE;
RX MEDLINE: 88030023.
RA Nakajima H., Noguchi T., Yamasaki T., Kono N., Tanaka T., Tarui S.;
RT "Cloning of human muscle phosphofructokinase cDNA."
RL FEBS Lett. 223:113-116(1987).
RN [4]
RP SEQUENCE OF 1-27 FROM N.A.
RX TISSUE=MUSCLE;
RX MEDLINE: 89306652.
RA Valdez B.C., Chen Z., Sosa M.G., Younathan E.S., Chang S.H.;
RT "Human 6-phosphofructo-1-kinase gene has an additional intron
RT upstream of start codon."
RL Gene 76:167-169(1989).
RN [5]
RP REVIEW ON GSD-VII VARIANTS.
RX MEDLINE: 96055509.
RA Raben N., Sherman J.B.;
RT "Mutations in muscle phosphofructokinase gene."
RL Hum. Mutat. 6:1-6(1995).
RN [6]
RP VARIANTS GSD-VII GLN-99; ASP-208 AND HIS-695.
RX MEDLINE: 95126102.
RA Raben N., Exelbert R., Spiegel R., Sherman J.B., Plotz P.,
RA Heinisch J.;
RT "Functional expression of human mutant phosphofructokinase in yeast:
RT genetic defects in French Canadian and Swiss patients with
RT phosphofructokinase deficiency."
RL Am. J. Hum. Genet. 56:131-141(1995).
CC -1- CATALYTIC ACTIVITY: ATP + D-FRUCTOSE 6-PHOSPHATE = ADP +
CC D-FRUCTOSE 1,6-BISPHOSPHATE.
CC -1- COFACTOR: REQUIRES MAGNESIUM.
CC -1- ENZYME REGULATION: ALLOSTERIC ENZYME ACTIVATED BY ADP, AMP, OR
CC FRUCTOSE BISPHOSPHATE AND INHIBITED BY ATP OR CITRATE.
CC -1- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
CC -1- SUBUNIT: TETRAMER. MUSCLE IS M4, LIVER IS L4, AND RED CELL IS M3L,
CC M2L2, OR M13.
CC -1- DISEASE: DEFECTS IN PFKM ARE THE CAUSE OF GLYCOGEN STORAGE
CC DISEASE VII (GSD-VII) (ALSO KNOWN AS TARUI'S DISEASE); A DISEASE
CC CHARACTERIZED BY EXERCISE INTOLERANCE WITH ASSOCIATED NAUSEA AND
CC VOMITING. SHORT BURSTS OF INTENSE ACTIVITY ARE PARTICULARLY
CC DIFFICULT. SEVERE MUSCLE CRAMPS AND MYOGLOBINURIA DEVELOP AFTER
CC VIGOROUS EXERCISE. MOST PATIENTS OBTAIN A "SECOND WIND" WHEN THE
CC ONSET OF EXERCISE IS FOLLOWED BY A BRIEF REST PERIOD. IN TIME

CC PATIENTS ADJUST THEIR ACTIVITY LEVEL AND ARE WELL COMPENSATED.
CC -|- MISCELLANEOUS: IN HUMAN PFK EXISTS AS A SYSTEM OF 3 TYPES OF
CC SUBUNITS, PFKM (MUSCLE), PFKL (LIVER) AND PFKP (PLATELET)
CC ISOENZYMES.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M59741; AAA82938.1; JOINED.
CC EMBL; M59720; AAA82938.1; JOINED.
CC EMBL; M59721; AAA82938.1; JOINED.
CC EMBL; M59722; AAA82938.1; JOINED.
CC EMBL; M59723; AAA82938.1; JOINED.
CC EMBL; M59724; AAA82938.1; JOINED.
CC EMBL; M59725; AAA82938.1; JOINED.
CC EMBL; M59726; AAA82938.1; JOINED.
CC EMBL; M59727; AAA82938.1; JOINED.
CC EMBL; M59728; AAA82938.1; JOINED.
CC EMBL; M59729; AAA82938.1; JOINED.
CC EMBL; M59730; AAA82938.1; JOINED.
CC EMBL; M59731; AAA82938.1; JOINED.
CC EMBL; M59732; AAA82938.1; JOINED.
CC EMBL; M59733; AAA82938.1; JOINED.
CC EMBL; M59734; AAA82938.1; JOINED.
CC EMBL; M59735; AAA82938.1; JOINED.
CC EMBL; M59736; AAA82938.1; JOINED.
CC EMBL; M59737; AAA82938.1; JOINED.
CC EMBL; M59738; AAA82938.1; JOINED.
CC EMBL; M59739; AAA82938.1; JOINED.
CC EMBL; M59740; AAA82938.1; JOINED.
CC EMBL; M26066; AAA60068.1; JOINED.
CC EMBL; Y00698; CAA60692.1; JOINED.
CC EMBL; M24925; AAA36436.1; JOINED.
CC PIR; PH0223; PH0223.
CC PIR; S00158; S00158.
CC PIR; JS0205; JS0205.
CC HSP; P00512; 4PFK.
CC HSC-2DPAGE; P08237; HUMAN.
CC MIM; 171850; -.
CC MIM; 232800; -.
CC PRINTS; PR00476; PHFCTKINASE.
CC PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 2.
CC PFAM; PF00365; PFK; 2.
CC Kinase: Transferase; Glycolysis; Repeat; Allosteric enzyme;
CC Phosphorylation; Magnesium; Multigene family; Disease mutation;
CC Glycogen storage disease.
CC INIT_MET 0
CC MOD_RES 774 774 PHOSPHORYLATION (BY SIMILARITY).
CC VARIANT 38 38 R -> L (IN GSD-VII; ASKENAZI).
CC VARIANT 38 38 /FTIG-VAR_006063.
CC VARIANT 38 38 R -> P (IN GSD-VII; ITALIAN).
CC VARIANT 99 99 R -> Q (IN GSD-VII; SWISS).
CC VARIANT 208 208 /FTIG-VAR_006064.
CC VARIANT 542 542 /FTIG-VAR_006065.
CC VARIANT 685 685 /FTIG-VAR_006066.
CC VARIANT 695 695 /FTIG-VAR_006067.
CC VARIANT 695 695 /FTIG-VAR_006068.
CC VARIANT 779 AA; 85051 MW; 4C9F384A7A5A5750 CRC64;
CC SEQUENCE 779 AA; 85051 MW; 4C9F384A7A5A5750 CRC64;
Query Match 43.1%; Score 62; DB 1; Length 779;
Best Local Similarity 50.0%; Pred. No. 4.74e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 123 GADTFRSEWSDL 134
|:|:|:|:|:|:|:
Qy 195 GPDTPSDWSEV 206

Search completed: Wed May 10 13:27:00 2000
Job time : 100 secs.

W A R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 13:27:18 2000; Maspar time 232.47 Seconds
Tabular output not generated. 5.667 Million cell updates/sec

Title: >US-09-376-430-2
Description: (188-206) from US09376430A.pep (18 of 25)
Perfect Score: 144
Sequence: 1 KAMEDVYGPDTYPSDWEV 19

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptembll2

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 28.889; Variance 45.018; scale 0.642

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	69	47.9	2532	3 Q9Y7D5	POLYKETIDE SYNTHASE.	9.58e+01
2	68	47.2	102	1 Q29936	HYPOTHETICAL 12.3 KD P	1.42e+00
3	68	47.2	229	2 Q87971	RESPONSE REGULATOR YUK	1.42e+00
4	67	46.5	676	1 Q9YEQ9	676AA LONG HYPOTHETICA	2.10e+00
5	67	46.5	4436	1 Q58659	4436AA LONG HYPOTHETIC	2.10e+00
6	66	45.8	247	2 Q85822	RESPONSE REGULATOR YSP	3.09e+00
7	66	45.8	255	2 Q9X207	CONSERVED HYPOTHETICAL	3.09e+00
8	66	45.8	328	9 Q92XA8	GP39.	3.09e+00
9	65	45.1	8243	5 Q96554	TYPE I FATTY ACID SYNT	4.54e+00
10	64	44.4	609	14 Q56942	E6, E7, E1, E2, E4, L2	6.64e+00
11	64	44.4	609	14 Q56942	E6, E7, E1, E2, E4, L2	6.64e+00
12	64	44.4	609	14 Q56942	E6, E7, E1, E2, E4, L2	6.64e+00
13	64	44.4	609	14 Q56942	E6, E7, E1, E2, E4, L2	6.64e+00
14	63	43.8	140	3 Q94341	PROTEIN HTEPC (ORF2 PR	6.64e+00
15	63	43.8	140	3 Q94341	PROTEIN HTEPC (ORF2 PR	6.64e+00
16	63	43.8	140	3 Q94341	PROTEIN HTEPC (ORF2 PR	6.64e+00
17	62	43.1	206	11 Q63736	VERY LOW DENSITY LIPOP	9.67e+00
18	62	43.1	206	11 Q63736	VERY LOW DENSITY LIPOP	9.67e+00
19	62	43.1	206	11 Q63736	VERY LOW DENSITY LIPOP	9.67e+00
20	61	42.4	118	2 Q92NP3	BETA-D-GALACTOSIDASE.	1.40e+01
					RAT PFK-M (FRAGMENT).	1.40e+01
					YOAD.	1.40e+01
					F44FL1.5 PROTEIN.	1.40e+01
					PHENOL HYDROXYLASE COM	2.03e+01

21	61	42.4	324	4	095268	ORIGIN RECOGNITION COM	2.03e+01
22	61	42.4	370	5	076633	H22K11.4 PROTEIN.	2.03e+01
23	61	42.4	388	3	074294	CYTCHROME P450 MONOOX	2.03e+01
24	61	42.4	435	11	09WUV0	ORC5-RELATED PROTEIN.	2.03e+01
25	61	42.4	446	5	Q09846	HYPOTHETICAL 51.7 KD P	2.03e+01
26	61	42.4	449	1	Q9Y9H2	449AA LONG HYPOTHETICA	2.03e+01
27	61	42.4	486	11	Q9WVE8	CYTOSOLIC PHOSPHOPRO	2.03e+01
28	61	42.4	568	10	Q23123	F19G10.5 PROTEIN.	2.03e+01
29	61	42.4	932	2	051674	PENICILLIN-BINDING PRO	2.03e+01
30	61	42.4	1878	14	Q66580	RNA SEQUENCE	2.03e+01
31	61	42.4	2493	14	Q88789	NONSTRUCTURAL POLYPROT	2.03e+01
32	60	41.7	110	5	Q16290	F32D1.8 PROTEIN.	2.92e+01
33	60	41.7	137	14	Q07344	STRUCTURAL PROTEIN P11	2.92e+01
34	60	41.7	275	3	Q14088	POTATIVE OXIDOREDUCTAS	2.92e+01
35	60	41.7	322	3	Q14834	URACIL-DNA GLYCOSYLASE	2.92e+01
36	60	41.7	335	5	Q17622	C04B4.3 PROTEIN.	2.92e+01
37	60	41.7	421	13	Q73739	ENDOTHELIN TYPE A RECE	2.92e+01
38	60	41.7	459	1	Q59265	459AA LONG HYPOTHETICA	2.92e+01
39	60	41.7	522	5	Q23593	ZK792.2 PROTEIN.	2.92e+01
40	60	41.7	1082	5	Q01905	SIMILARITY TO HUMAN SK	2.92e+01
41	60	41.7	1187	5	Q93254	C23H4.3 PROTEIN.	2.92e+01
42	59	41.0	160	14	Q89715	NATURAL KILLER CELL RE	4.18e+01
43	59	41.0	188	14	Q57241	NK CELL RECEPTOR HOMOL	4.18e+01
44	59	41.0	333	2	Q87541	C-TYPE CYTOCHROME PREC	4.18e+01
45	59	41.0	1657	5	P90761	C27H6.1 PROTEIN.	4.18e+01

ALIGNMENTS

RESULT 1
ID Q9Y7D5 PRELIMINARY; PRT; 2532 AA.
AC Q9Y7D5;
DT 01-NOV-1999 (TRENBLREL. 12, Created)
DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLREL. 12, Last annotation update)
DE POLYKETIDE SYNTHASE.
GN LOVF.

OS Aspergillus terreus.
OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC20542;
RA KENNEDY J., AUCLAIR K., KENDREW S.G., PARK C., VEDERAS J.C.,
RA HUTCHINSON C.R.;
RT "Accessory Proteins Modulate Polyketide Synthase Activity During
RT Lovastatin Biosynthesis."
RL Science 0:0-0(1999).
DR EMBL; AF141925; AAD34559.1;
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
KW Transferase.
SQ SEQUENCE 2532 AA; 276638 MW; D6E9F44D CRC32;

Query Match 47.98; Score 69; DB 3; Length 2532;
Best Local Similarity 53.3%; Pred. No. 9.58e+01;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 1838 DTVEPKLPDDWVEI 1852
QY 192 DVYGPDTYPSDWEV 206

RESULT 2
ID Q29936 PRELIMINARY; PRT; 102 AA.
AC Q29936;
DT 01-JAN-1998 (TRENBLREL. 05, Created)
DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)
DT 01-AUG-1998 (TRENBLREL. 07, Last annotation update)
DE HYPOTHETICAL 12.3 KD PROTEIN.
GN AF0308.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.

```

RN RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAUGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBECK R., GOICAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., RAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESSE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL: AE001083; AAB90926.1; -.
DR TIGR: AF0308; -.
KW Hypothetical protein.
SQ SEQUENCE 102 AA; 12265 MW; AA370DE2 CRC32;

Query Match 47.2%; Score 68; DB 1; Length 102;
Best Local Similarity 58.3%; Pred. No. 1.42e+00;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 29 DIFIPDEYPKDW 40
QY 192 DVYGPDTYPSDW 203
I:::|||||

RESULT 3
ID O87971; PRELIMINARY; PRT; 229 AA.
AC O87971;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE RESPONSE REGULATOR YUKR (FRAGMENT).
GN YUKR.
OS Yersinia ruckeri.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1315;
RA ATKINSON S., THROUP J.P., WILLIAMS P., STEWART G.S.A.B.;
RT "A hierarchical quorum sensing system in Yersinia pseudotuberculosis
RT is involved in the regulation of motility and morphology.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DDJB databases.
DR EMBL: AF079135; AAC28702.1; -.
DR PFAM: PF00196; GerE; 1.
FT NON_TER 1
SQ SEQUENCE 229 AA; 26142 MW; 2743EFAF CRC32;

Query Match 47.2%; Score 68; DB 2; Length 229;
Best Local Similarity 33.3%; Pred. No. 1.42e+00;
Matches 5; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Db 25 DVFIYTSYDPEWADI 39
QY 192 DVYGPDTYPSDSEV 206
I:::|||||

RESULT 4
ID Q9YEQ9 PRELIMINARY; PRT; 676 AA.
AC Q9YEQ9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE 676AA LONG HYPOTHETICAL CLEAVAGE AND POLYADENYLATION FACTOR SUBUNIT.
GN APE0522.
OS Aeropyrum pernix.

```

```

OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K1;
RC MEDLINE: 99310339.
RA KAWARABAYASHI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOVAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL: AP000059; BAA79487.3; -.
SQ SEQUENCE 676 AA; 76444 MW; ECCDF8A CRC32;

Query Match 46.5%; Score 67; DB 1; Length 676;
Best Local Similarity 72.7%; Pred. No. 2.10e+00;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 233 YGPDSPYFWS 243
QY 194 YGPDTPSDWS 204
I:::|||||

RESULT 5
ID O58659 PRELIMINARY; PRT; 4436 AA.
AC O58659;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE 4436AA LONG HYPOTHETICAL PROTEIN.
GN PH0954.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE: 98344137.
RA KAWARABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOVAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL: AP000004; BAA30051.1; -.
DR PFAM: PF00041; fn3; 1.
DR PFAM: PF00801; PKD; 2.
SQ SEQUENCE 4436 AA; 497015 MW; 2D9608CA CRC32;

Query Match 46.5%; Score 67; DB 1; Length 4436;
Best Local Similarity 41.7%; Pred. No. 2.10e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 1113 IYSTNAYPRNWS 1124
QY 193 YGPDTPSDWS 204
I:::|||||

RESULT 6
ID O85822 PRELIMINARY; PRT; 247 AA.
AC O85822;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE RESPONSE REGULATOR YSPR.
GN YSPR.
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

```

OC Viruses: dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC Lambda phage group.
RN [1]
RN RC SEQUENCE FROM N.A.
RN RC STRAIN-NORWICH STOCK;
RA SMITH M.C.M.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RN RC SEQUENCE FROM N.A.
RN RC STRAIN-NORWICH STOCK;
RX MEDLINE: 99162580.
RA HENDRIX R.W., SMITH M.C.M., BURNS N., FORD M.E., HATFULL G.F.;
RT "Evolutionary relationships among diverse bacteriophages and
RT prophages: all the world's a phage.";
RT Proc. Natl. Acad. Sci. U.S.A. 96:2192-2197(1999).
DR EMBL: AJ006589; CAA07109.1; -.
SQ SEQUENCE 328 AA; 34843 MW; 4772C36D CRC32;

Query Match 45.88; Score 66; DB 9; Length 328;
Best Local Similarity 40.08; Pred. No. 3.09e+00;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 168 VEPTYPAEAYPYDWT 182
:| |::|||:
QY 190 MEDVYGDPDTPSDWS 204

RESULT 9
ID ID O96554 PRELIMINARY; PRT; 8243 AA.
AC O96554;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE TYPE I FATTY ACID SYNTHASE.
GN FAS1.
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
RN [1]
RN RC SEQUENCE FROM N.A.
RN RC STRAIN=KSU-1;
RA ZHU G., MARCHENKA M.J., WOODS K.M., UPTON S.J., KEITHLY J.S.;
RT "Characterization of a type I FAS gene in the parasitic protozoan
RT Cryptosporidium parvum.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF082993; AAC99407.1; -.
DR HSP: P28304; 1QOR.
SQ SEQUENCE 8243 AA; 920806 MW; 0FC76B0F CRC32;

Query Match 45.18; Score 65; DB 5; Length 8243;
Best Local Similarity 46.78; Pred. No. 4.54e+00;
Matches 7; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Db 4975 MESVIAPKVY-GAWN 4988
|||::|||:
QY 190 MEDVYGDPDTPSDWS 204

RESULT 10
ID ID O56942 PRELIMINARY; PRT; 609 AA.
AC O56942;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE E5, E7, E1, E2, E4, L2, AND L1 GENES.
GN E1.
OS Human papillomavirus type 76.
OC Viruses: dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
RN [1]
RN RC SEQUENCE FROM N.A.
RX MEDLINE: 98118461.
RA DELIUS H., SAEGLING B., BERGMANN K., SHAMANIN V., DE VILLIERS E.M.;
RT "The genomes of three of four novel HPV types, defined by differences

```

RT of their L1 genes, show high conservation of the E7 gene and the
 RT URR.;"
 RL Virology 240:359-365(1998).
 DR EMBL; Y15174; CAA75458.1; -.
 DR PFAM; PF00519; E1; 1.
 DR PFAM; PF00524; E1_N; 1.
 SQ SEQUENCE 609 AA; 70111 MW; 63A50A27 CRC32;

Query Match 44.4%; Score 64; DB 14; Length 609;
 Best Local Similarity 50.0%; Pred. No. 6.64e+00;
 Matches 8; Conservative 3; Mismatches 3; Indels 2; Gaps 2;

Db 283 SMDSSVYAHGTYP-DW 297
 QY 189 AMED-VYGPDTYPSDW 203

RESULT 11
 ID O56936 PRELIMINARY; PRT; 609 AA.
 AC O56936;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE E6, E7, E1, E2, E4, L2, AND L1 GENES.
 GN E1.
 OS Human papillomavirus type 75.
 OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 98118461.
 RA DELIUS H., SAEGLING B., BERGMANN K., SHAMANIN V., DE VILLIERS E.M.;
 RT "The genomes of three of four novel HPV types, defined by differences
 of their L1 genes, show high conservation of the E7 gene and the
 URR.;"
 RL Virology 240:359-365(1998).
 DR EMBL; Y15173; CAA75451.1; -.
 DR PFAM; PF00519; E1; 1.
 DR PFAM; PF00524; E1_N; 1.
 SQ SEQUENCE 609 AA; 70276 MW; 3F2B3DB9 CRC32;

Query Match 44.4%; Score 64; DB 14; Length 609;
 Best Local Similarity 50.0%; Pred. No. 6.64e+00;
 Matches 8; Conservative 3; Mismatches 3; Indels 2; Gaps 2;

Db 283 SMDSSVYAHGTYP-DW 297
 QY 189 AMED-VYGPDTYPSDW 203

RESULT 12
 ID P70222 PRELIMINARY; PRT; 677 AA.
 AC P70222; P70221;
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE PROTEIN HTF9C (ORF2 PRODUCT).
 GN HTF9C OR HTF9-C.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN SEQUENCE FROM N.A.
 RP STRAIN-C57 BLACK;
 RC MEDLINE; 91365246.
 RA BRESSAN A., SOMMA M.P., LEWIS J., SANTOLMAZZA C., COPELAND N.G.,
 RA GILBERT D.J., JENKINS N.A., LAVIA P.;
 RT "Characterization of the opposite-strand genes from the mouse
 bidirectionally transcribed Htf9 locus.;"
 RL Gene 103:201-209(1991).
 RN SEQUENCE FROM N.A.
 RP STRAIN-C57 BLACK;
 RC MEDLINE; 93281396.

RA STAPLETON G., SOMMA M.P., LAVIA P.;
 RT "Cell type-specific interactions of transcription factors with a
 RT housekeeping promoter in vivo.;"
 RL Nucleic Acids Res. 21:2465-2471(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57 BLACK;
 RA LAVIA P.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57 BLACK;
 RA GUARGUAGLINI G.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57 BLACK;
 RX MEDLINE; 97344268.
 RA GUARGUAGLINI G., BATTISTONI A., PITTOGGI C., DI MATTEO G.,
 RA DI FIORE B., LAVIA P.;
 RT "Expression of the murine RanBp1 and Htf9-c genes is regulated from a
 RT shared bidirectional promoter during cell cycle progression.;"
 RL Biochem. J. 325:277-286(1997).
 RN [6]
 RP SEQUENCE OF 1-277 FROM N.A.
 RX MEDLINE; 88054976.
 RA LAVIA P., MACLEOD D., BIRD A.;
 RT "Coincident start sites for divergent transcripts at a randomly
 RT selected CpG-rich island of mouse.;"
 RL EMBO J. 6:2773-2779(1987).
 RN [7]
 RP SEQUENCE OF 1-277 FROM N.A.
 RX MEDLINE; 92349415.
 RA TYNDALL C., WATT F., MOLLOY P.L., VINCENT P.C., FROMMER M.;
 RT "Binding of proteins from embryonic and differentiated cells to a
 RT bidirectional promoter contained within a CpG island.;"
 RL J. Mol. Biol. 226:289-299(1992).
 RN [8]
 RP SEQUENCE OF 1-277 FROM N.A.
 RX MEDLINE; 96114085.
 RA DI MATTEO G., FUSCHI P., ZERFASS K., MORETTI S., RICORDY R.,
 RA GENCIARELLI C., TRIPODI M., JANSEN-DURR P., LAVIA P.;
 RT "Transcriptional control of the Htf9-A/RanBp-1 gene during the cell
 RT cycle.;"
 RL Cell Growth Differ. 6:1213-1224(1995).
 DR EMBL; X56044; CAA39515.1; -.
 DR EMBL; X05830; CAA29277.1; -.
 DR MGD; MGI:96270; Htf9C.
 FT CHAIN 1 677 PROTEIN HTF9C.
 SQ SEQUENCE 677 AA; 74870 MW; 68099DC9 CRC32;

Query Match 44.4%; Score 64; DB 11; Length 677;
 Best Local Similarity 38.5%; Pred. No. 6.64e+00;
 Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 358 YDPETYGHKQL 370
 QY 194 YGPDYPSDSEV 206

RESULT 13
 ID O77505 PRELIMINARY; PRT; 845 AA.
 AC O77505;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE VERY LOW DENSITY LIPOPROTEIN RECEPTOR VLDL-R2.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE: 99018030.
RA MAGRANE J., REINA M., PAGAN R., LUNA A., CASAROLI-MARANO R.P.,
RA ANGELIN B., GAFVELS M., VILARO S.;
RT "Bovine aortic endothelial cells express a variant of the very low
RT density lipoprotein receptor that lacks the O-linked sugar domain."
RL J. Lipid Res. 39:2172-2181(1998).
DR EMBL: AF016537; AD03484.1; -
DR HSSP: P01130; 1AJJ.
DR DR PROSITE; PS01187; EGF_CA; 1.
DR DR PROSITE; PS01209; LDLRA_1; 8.
DR PFAM: PF00057; ldl_recept_a; 8.
DR PFAM: PF00058; ldl_recept_b; 5.
DR PRINTS; PR00261; LDLRECEPT.
KW Lipoprotein; Glycoprotein; EGF-like domain.
SQ SEQUENCE 845 AA; 95298 MW; 02275B7B CRC32;

Query Match 44.4%; Score 64; DB 6; Length 845;
Best Local Similarity 37.5%; Pred. No. 6.64e+00;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 507 KMDNVNPAIAVDW 522
| : : : | | : : |
QY 188 KAMEDVYGPDTYPSDW 203

RESULT 14
ID O94341 PRELIMINARY; PRT; 140 AA.
AC O94341;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE VERY HYPOTHETICAL PROTEIN.
GN SPBC1271.08C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-;
RA LYNE M., RAJANDREAM M.A., BARRELL B.G., BECK A., BORZYM K., KLAGES S.,
RA LANGER I., REINHARDT R;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL034353; CAA22198.1; -
SQ SEQUENCE 140 AA; 16689 MW; 17D8D3F0 CRC32;

Query Match 43.8%; Score 63; DB 3; Length 140;
Best Local Similarity 43.8%; Pred. No. 9.67e+00;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 48 NDTYSSDTYELDWYDL 63
: | : : | | | | : : :
QY 191 EDVYGPDTYPSDWSEV 206

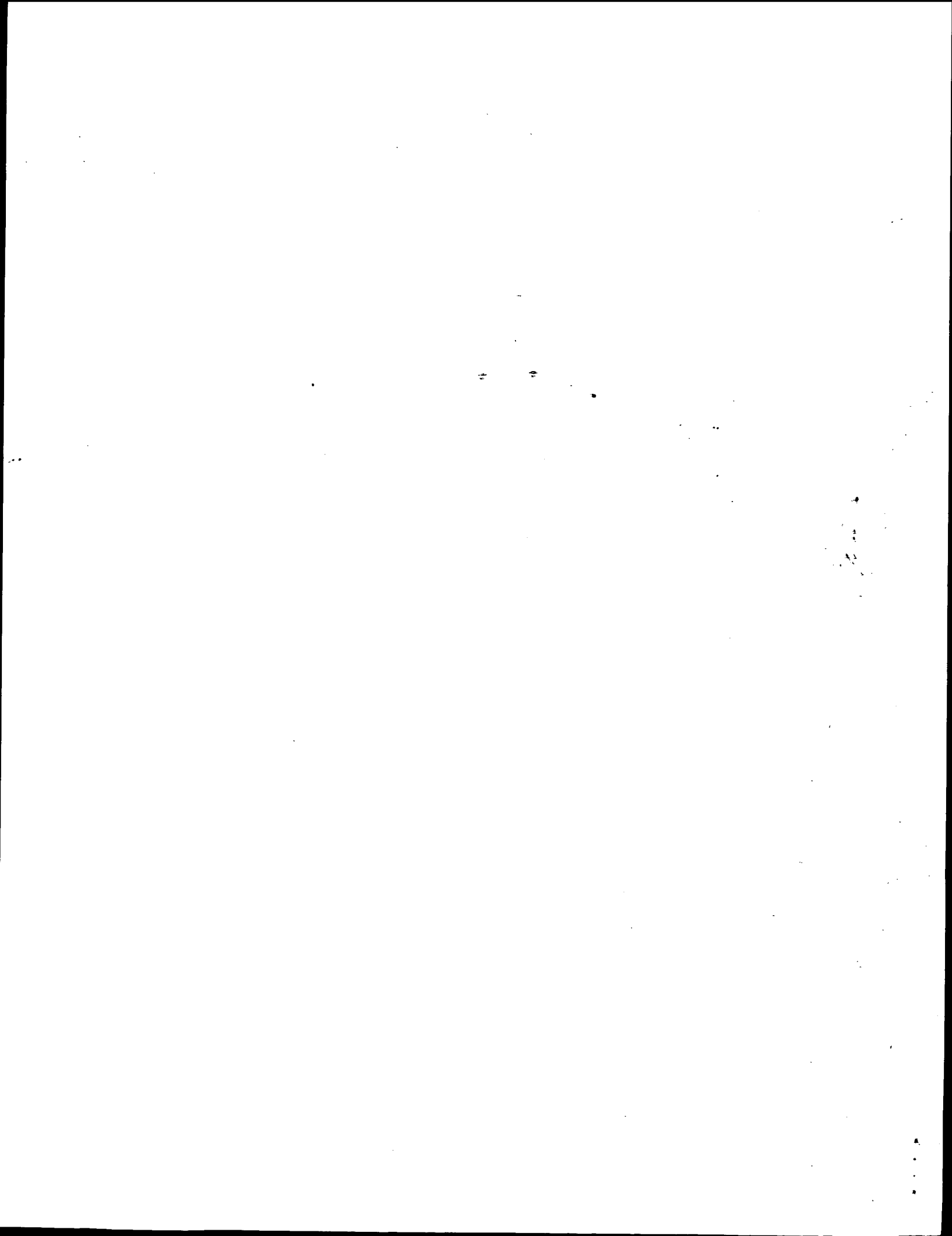
RESULT 15
ID Q40238 PRELIMINARY; PRT; 157 AA.
AC Q40238;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE HOMEBOX TRANSCRIPTION FACTOR HOX7 (FRAGMENT).
GN HOX7.
OS Lycopersicon peruvianum (Peruvian tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
OC Solanum.
RN [1]
RP SEQUENCE FROM N.A.
RA SCHARF K.D., GANAL M.;
RL Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL: X67602; CAA47871.1; -

DR PFAM: PF00046; homeobox; 1.
KW Homeobox; DNA-binding; Nuclear protein.
FT NON_TER 1
FT NON_TER 157
SQ SEQUENCE 157 AA; 17403 MW; E5A6C719 CRC32;

Query Match 43.8%; Score 63; DB 10; Length 157;
Best Local Similarity 50.0%; Pred. No. 9.67e+00;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 26 TLERVYAMETYPSE 39
: | : : | | : : |
QY 189 AMEDVYGPDTYPSD 202

Search completed: Wed May 10 13:31:22 2000
Job time : 244 secs.



M P S R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 13:40:30 2000; MasPar time 3.32 Seconds
Tabular output not generated. 171.130 Million cell updates/sec

Title: >US-09-376-430-2
Description: (208-231) from US09376430A.pep (19 of 25)
Perfect Score: 184
Sequence: 1 CQORGEIRDACAETPTPPKPKLSK 24

Scoring table: PAM 150
Gap 11

Searched: 189963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:geneseq

Statistics: Mean 21.797; Variance 69.304; scale 0.315

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	90	48.9	572	1 Y04954	Mycobacterium species	6.92e-02
2	67	36.4	610	1 W29749	Tumor suppressor prote	1.85e-01
3	66	35.9	763	1 W31852	Mycobacterium tubercul	2.33e-01
4	64	34.8	31	1 W21999	Insulin C chain peptid	3.69e-01
5	64	34.8	501	1 R70996	Human alpha-1A adrener	3.69e-01
6	64	34.8	501	1 R90040	Human alpha-1A adrener	3.69e-01
7	62	33.7	114	1 Y11113	H. pylori ORF hpel1036	5.82e-01
8	61	33.2	384	1 W83354	Vernonia galamensis	7.28e-01
9	61	33.2	778	1 R94900	Human tastin.	7.28e-01
10	60	32.6	771	1 R14256	Human GCSF receptor en	9.11e-01
11	59	32.1	231	1 W20982	H. pylori secreted or	1.14e-02
12	59	32.1	1104	1 R24305	NF-X1 DNA-binding prot	1.14e-02
13	58	31.5	36	1 P50075	Peptide with alpha-amy	1.42e-02
14	58	31.5	919	1 W18580	Potato alpha-glucosida	1.42e-02
15	58	31.5	1061	1 W87504	Human N-methyl-D-aspar	1.42e-02
16	58	31.5	1212	1 W87503	Human N-methyl-D-aspar	1.42e-02
17	58	31.5	2213	1 W26356	Rabbit LDL receptor an	1.42e-02
18	58	31.5	2408	1 R24305	Translation of ORF 1 c	1.42e-02
19	57	31.0	256	1 Y00923	Alternatively spliced	1.77e-02
20	57	31.0	276	1 Y00922	Human CLAR1 protein se	1.77e-02
21	57	31.0	472	1 W37940	Amino acid sequence of	1.77e-02
22	57	31.0	872	1 Y07110	Colon cancer associate	1.77e-02
23	57	31.0	1045	1 W57448	M. tuberculosis isoleu	1.77e-02

24	57	31.0	1052	1 W37412	Human G-protein couple	1.77e-02
25	57	31.0	1572	1 W99301	Human BAI2 protein.	1.77e-02
26	57	31.0	214	1 W26357	Human LDL receptor ana	1.77e-02
27	56	30.4	121	1 W23892	Murine fibroblast grow	2.20e-02
28	56	30.4	183	1 P60623	Sequence A encoded by	2.20e-02
29	56	30.4	284	1 W98412	H. pylori GPO 269 pro	2.20e-02
30	56	30.4	372	1 R42310	Lignin peroxidase.	2.20e-02
31	56	30.4	372	1 R42891	Lignin peroxidase.	2.20e-02
32	56	30.4	387	1 W94461	Human Sox1 protein.	2.20e-02
33	56	30.4	469	1 W37061	HIV-1 breakthrough iso	2.20e-02
34	56	30.4	474	1 W15253	Human brain-specific a	2.20e-02
35	56	30.4	481	1 W80811	The amino acid sequenc	2.20e-02
36	56	30.4	494	1 W37060	HIV-1 breakthrough iso	2.20e-02
37	56	30.4	527	1 R63446	Large subunit of ADP-g	2.20e-02
38	56	30.4	594	1 W15256	Human brain-specific a	2.20e-02
39	55	29.9	235	1 R57089	Fasciola hepatica larv	2.73e-02
40	55	29.9	319	1 W18658	Fragmented human NF-L	2.73e-02
41	55	29.9	362	1 P81100	Sequence of RI fragmen	2.73e-02
42	55	29.9	481	1 Y04145	Rat Tango-76 protein.	2.73e-02
43	55	29.9	541	1 W12695	G-protein parathyroid	2.73e-02
44	55	29.9	610	1 R10923	Maize protein encoded	2.73e-02
45	55	29.9	5405	1 W14749	IGG-Fc binding protein	2.73e-02

ALIGNMENTS

RESULT 1
ID Y04954 standard; Protein; 572 AA.
AC Y04954;
DT 06-JUL-1999 (first entry)
DE Mycobacterium species protein sequence 41#2.
KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection.
OS Mycobacterium sp.
PN WO9909186-A2.
PD 25-FEB-1999.
PF 14-AUG-1998; F01813.
PR 11-SEP-1997; FR-011325.
PR 14-AUG-1997; FR-010404.
PA (INSP) INST PASTEUR.
PI Gicquel B, Lim EM, Pellicic V, Portnoi D, Goguet de la Salmoniere Y,
PI Guigueno A;
DR WPI; 99-181045/15.
DR N-PSDB; X34206.
PT Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in
PT infection-associated protein expression
PS Claim 32; Fig 41P; 309pp; French.
CC Sequences Y04742-Y05000 and Y07201-Y07204 represent secreted proteins
CC from various Mycobacterium species microorganisms. The encoding
CC nucleotide sequences can be used as primers and probes for methods
CC for detecting and identifying mycobacteria, especially belonging to
CC the M. tuberculosis complex. The encoded proteins can be used in
CC vaccines for immunisation against a bacterial or viral infection.
SQ Sequence 572 AA;

Query Match 48.9%; Score 90; DB 1; Length 572;
Best Local Similarity 43.5%; Pred. No. 6.92e-02;
Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Db	518	CWHREAIWXRCLXAPXPKTWLS	540
Qy	208	CWQORGEIRDACAETPTPPKPKLS	230

RESULT 2
ID W29749 standard; Protein; 610 AA.
AC W29749;
DT 11-JAN-1999 (first entry)
DE Tumor suppressor protein menin.
KW Familial multiple endocrine neoplasia type 1; FMEN1; MEN1;
KW menin; tumour suppressor gene; cancer; marker; diagnosis;
KW gene therapy; human.

OS Homo sapiens.
 PN WO9839439-A1.
 PD 11-SEP-1998.
 PF 04-MAR-1998; U04258.
 PR 05-MAR-1997; US-040269.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Agarwal SK, Burns AL, Chandrasekharappa SC, Collins FS,
 PI Dubelensky IV, Emmert-Buck MR, Guro SC, Liotta LA,
 PI Lubensky IA, Vanickam P, Marx SJ, Spiegel AM;
 DR WPI: 98-506360/43.
 DR N-PSDB: V57415-16.
 PT Protein and gene associated with multiple endocrine neoplasia type 1
 PT - useful in gene therapy and to diagnose sufferers of, and those
 PT susceptible to, this condition by detecting protein absence or gene
 PT mutation(s)
 PS Claim 11; Page 56-57; 75pp; English.
 CC This is the amino acid sequence of menin, a 67.5 kDa polypeptide
 CC encoded by a novel human tumour suppressor gene, MEN1 (see V57416),
 CC that is associated with multiple endocrine neoplasia type 1. Menin
 CC exhibits no apparent amino acid sequence similarities to previously
 CC known proteins. The lack of a functional menin polypeptide, either
 CC by absence of the protein, its alteration and/or associated
 CC mutations in the MEN1 gene, have been identified in individuals
 CC with familial multiple endocrine neoplasia type 1 (FMEN1) and
 CC suffering from multiple endocrine neoplasia type 1. The
 CC identification of MEN1 provides a new window into the mechanism
 CC of endocrine tumorigenesis, facilitates accurate early diagnosis of
 CC MEN1 associated cancers, and provides preclinical identification of
 CC individuals with the FMEN1 syndrome, i.e. cancer-free individuals
 CC that are at high risk of acquiring MEN1 associated tumours. It
 CC also provides a basis for gene therapy. Menin can be expressed in
 CC eukaryotic or prokaryotic host cells and used to identify menin
 CC specific reagents (e.g. antibodies) useful for detecting menin in
 CC humans cells or tissues, especially from patients suspected of
 CC being at risk from multiple endocrine neoplasia type 1.
 CC Sequence 610 AA;
 SQ

Query Match 36.4%; Score 67; DB 1; Length 610;
 Best Local Similarity 45.8%; Pred. No. 1.85e+01;
 Matches 11; Conservative 5; Mismatches 7; Indels 1; Gaps 1;
 DB 479 RRGPRRESPEPPPPKPKALDK 502
 QY 209 WQGEIRDACAETPTPK-PKLSK 231

RESULT 3
 ID W31852 standard; Protein; 763 AA.
 AC W31852;
 DT 27-APR-1998 (first entry)
 DE Mycobacterium tuberculosis 74 kDa protein.
 KW Tuberculosis; mycobacteria; infection; diagnosis;
 KW antimycobacterial; antibiotic; vaccine.
 OS Mycobacterium tuberculosis.
 PN WO9741252-A2.
 PD 06-NOV-1997.
 PF 18-APR-1997; E01973.
 PR 29-APR-1996; DE-017184.
 PA (GBFB) GBF GES BIOTECH FORSCHUNG GMBH.
 PI Espitia C, Honisch C, Moreno C, Singh M;
 DR WPI: 97-549750/50.
 DR N-PSDB: T93610.
 PT New DNA and related proteins or RNA derived from M. tuberculosis -
 PT used for diagnosis of mycobacterial infections, monitoring
 PT vaccination and development of anti-mycobacterial agents
 PS Claim 5; Fig 13; 59pp; English.
 CC This novel 74 kDa protein is encoded by an open reading frame of
 CC a Mycobacterium tuberculosis DNA fragment (see T93610) containing
 CC polymorphic GC-rich sequences. Its amino acid sequence shows
 CC a high proline content, but there is no homology to any known
 CC proline-rich antigens of mycobacteria. Novel M. tuberculosis
 CC proteins (see W31851-57) are claimed. These can be produced as
 CC recombinant proteins, especially in bacterial, yeast, fungal or

CC higher eukaryote host cells, and used for diagnosing tuberculosis
 CC and other mycobacterial infections in humans or animals. The
 CC claimed proteins can also be used for epidemiological studies, for
 CC monitoring vaccination, and for the development of vaccines and
 CC anti-mycobacterial drugs.
 SQ Sequence 763 AA;
 Query Match 35.9%; Score 66; DB 1; Length 763;
 Best Local Similarity 40.9%; Pred. No. 2.33e+01;
 Matches 9; Conservative 4; Mismatches 8; Indels 1; Gaps 1;
 DB 38 CWMLVSAAPPCPPAPPAPPKPK 59
 QY 208 CWQGEIRDACAETP-TPPKPK 228

RESULT 4
 ID W21999 standard; peptide; 31 AA.
 AC W21999;
 DT 04-MAR-1998 (first entry)
 DE Insulin C chain peptide linker E.coli-N.
 KW Insulin C chain; linker; recombinant; fused protein; chimeric;
 KW interleukin 12; IL-12; gene therapy; mammalian host.
 OS Synthetic.
 PN WO9723639-A1.
 PD 03-JUL-1997.
 PF 20-DEC-1996; J03742.
 PR 22-DEC-1995; JP-350364.
 PR 22-DEC-1995; JP-334875.
 PA (TORA) TORAY IND INC.
 PI Akiyama H, Fujiwara H, Kanai S, Nakatani I, Tomura M;
 DR WPI: 97-351074/32.
 DR N-PSDB: T74633.
 PT Production of recombinant biologically active fused proteins - by
 PT culture of a host transformed with DNA encoding units of the fused
 PT protein joined by a linker removable during cellular processing
 PS Claim 4; Page 33; 77pp; Japanese.
 CC A method has been produced for producing biologically active fusion
 CC proteins by transforming a suitable host (preferably mammalian cells)
 CC with an expression vector containing DNA encoding a protein: A-X-B,
 CC where A and B = subunits of a dimeric protein or are each a
 CC biologically active protein; X is a linker polypeptide. The present
 CC sequence represents a specifically claimed example of X. The method
 CC can be used for the production of interleukin-12 (IL-12) using DNA
 CC coding for the 40 kD and 35 kD subunits of IL-12, joined by a suitable
 CC linker. The method is used for the production of biologically active
 CC fusion proteins for use as drugs, diagnostic agents or in gene therapy.
 CC IL-12 and IL-2 may be used in the treatment of rheumatoid arthritis,
 CC diabetes, cancer and viral infections.
 CC Sequence 31 AA;
 Query Match 34.8%; Score 64; DB 1; Length 31;
 Best Local Similarity 36.0%; Pred. No. 3.69e+01;
 Matches 9; Conservative 4; Mismatches 11; Indels 1; Gaps 1;
 DB 4 CWRPVGRIRHLCRISQSMFDATLSR 28
 QY 208 CWQR-GEIRDACAETPTPKPKLSK 231

RESULT 5
 ID R70996 standard; Protein; 501 AA.
 AC R70996;
 DT 08-JUN-1995 (first entry)
 DE Human alpha-1A adrenergic receptor.
 KW Human; alpha-1A adrenergic receptor; tissue specific receptor expression;
 KW reverse transcriptase-polymerase chain reaction; RT-PCR; heart mRNA;
 KW transmembrane region; primer; open reading frame; bovine; hippocampus;
 KW alpha-1C receptor; library; mammalian cell line; ligand; truncation;
 KW function; signal transduction signal; benign prostatic hypertrophy;
 KW flasteride; intraocular pressure; arrhythmia; CNS; prostatic cancer;
 KW male pattern baldness; seborrhea; female hirsutism; prostatitis; acne;

KW potassium channel opener; minoxidil; G-protein; ligand binding site.
 OS Homo sapiens.
 PN W09421560-A.
 PD 29-SEP-1994.
 PF 10-MAR-1994; U02609.
 PR 15-MAR-1993; US-032849.
 PA (MERI) MERCK & CO INC.
 PI Bayne ML, Clineschmidt BV, Strader CD;
 DR WPI: 94-316922/39.
 DR N-PSDB: Q72231.
 PT New DNA encoding alpha 1C and 1A human adrenergic receptors -
 PT used to detect selective inhibitors of alpha 1C, useful for
 PT treatment of benign prostatic hypertrophy
 PS Example 7; Fig 11; 120pp; English.
 CC This sequence represents a human alpha-1A adrenergic receptor. This
 CC protein is related to the alpha-1C adrenergic receptor of the invention
 CC which comprises an intact G-protein and ligand binding site. The
 CC cloned human alpha-1C receptor, when expressed in mammalian cell lines,
 CC is used to discover ligands that bind to the receptor and alter its
 CC function. Truncation at the 3' end of the sequence does not affect
 CC the functioning of the receptor, as long as the ligand binding domain
 CC and signal transduction segments of the receptor (G-protein interaction)
 CC are intact. Cells expressing the 1C receptor are used to identify
 CC compounds which bind to this receptor with at least a 12 fold greater
 CC affinity than to other adrenergic receptors. Compounds identified this
 CC way, are useful for treating the acute symptoms of benign prostatic
 CC hypertrophy, particularly when used in conjunction with finasteride.
 CC They may also be used to reduce intraocular pressure, control arrhythmia
 CC and other CNS events related to the 1C receptors, to treat or prevent
 CC male pattern baldness, seborrhea, female hirsutism, prostatitis,
 CC prostatic cancer and acne. For treating baldness they may be formulated
 CC with a potassium channel opener such as minoxidil. The cloned receptors
 CC can be used to screen for tissue specific receptor expression.
 SQ Sequence 501 AA;

Query Match 34.8%; Score 64; DB 1; Length 501;
 Best Local Similarity 31.8%; Pred. No. 3.69e+01;
 Matches 7; Conservative 6; Mismatches 8; Indels 1; Gaps 1;

Db 457 CWGRSGDRPSCAPKSPACRTR 478
 QY 208 CWQR-GEIRDACAETPTPKPK 228

RESULT 6
 ID R90040 standard; Protein; 501 AA.
 AC R90040;
 DT 27-OCT-1996 (first entry)
 DE Human alpha-1A adrenergic receptor protein.
 KW Alpha-1A adrenergic receptor; benign prostate hyperplasia;
 KW benign prostate hypertrophy.
 OS Homo sapiens.
 PN W09528397-A1.
 PD 26-OCT-1995.
 PF 13-APR-1995; U04590.
 PR 14-APR-1994; US-229276.
 PA (MERI) MERCK & CO INC.
 PI Huff JR, Lee H, Nerenberg JB, Thompson WJ;
 DR WPI: 93-373755/48.
 DR N-PSDB: T11605.
 PT New saccharin type derivs. - useful as selective alpha 1C adrenergic
 PT receptor antagonists for treating e.g. benign prostatic hyperplasia
 PS Disclosure; Page 134-135; 210pp; English.
 CC Compounds which are used to antagonise the receptor encoded by
 CC this sequence may be used for the treatment of urinary obstruction
 CC caused by benign prostatic hypertrophy.
 SQ Sequence 501 AA;

Query Match 34.8%; Score 64; DB 1; Length 501;
 Best Local Similarity 31.8%; Pred. No. 3.69e+01;
 Matches 7; Conservative 6; Mismatches 8; Indels 1; Gaps 1;

Db 457 CWGRSGDRPSCAPKSPACRTR 478

QY 208 CWQR-GEIRDACAETPTPKPK 228

RESULT 7
 ID Y11113 standard; Protein; 114 AA.
 AC Y11113;
 DT 08-JUN-1999 (first entry)
 DE H. pylori ORF hp610363_30517031_f3_3 secreted protein.
 KW Vaccine; probe; diagnostic; ORF; cell envelope protein;
 KW secreted protein; cytoplasmic protein; cellular protein.
 OS Helicobacter pylori.
 PN W09824475-A1.
 PD 11-JUN-1998.
 PF 05-DEC-1997; U22104.
 PR 14-JUL-1997; US-891928.
 PR 05-DEC-1996; US-759625.
 PR 25-MAR-1997; US-823745.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;
 DR WPI: 98-333051/29.
 DR N-PSDB: X30642.
 PT New isolated Helicobacter pylori nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of infection by
 PT H. pylori and other Helicobacter species
 PS Claims 37, 41; Page 278-279; 339pp; English.
 CC Recombinant or substantially pure preparations of H. pylori polypeptides
 CC are disclosed, together with the nucleic acids encoding them. In all,
 CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.
 CC Vaccines containing the nucleic acids or proteins are claimed, as are
 CC probes containing at least 8 nucleotides from the nucleic acid
 CC sequences. The vaccines are useful for treating or reducing the risk of
 CC H. pylori infections, and the probes can be used diagnostically for
 CC detecting the presence of Helicobacter in a sample. The products are
 CC also of use in screening for compounds having the ability to interfere
 CC with the H. pylori life cycle or to inhibit H. pylori infection.
 SQ Sequence 114 AA;

Query Match 33.7%; Score 62; DB 1; Length 114;
 Best Local Similarity 38.1%; Pred. No. 5.82e+01;
 Matches 8; Conservative 5; Mismatches 6; Indels 2; Gaps 2;

Db 44 CWRSATWRSARCSPPFRAPPR 64
 QY 208 CWQRGEI-RDA-CAETPTPK 226

RESULT 8
 ID W83354 standard; Protein; 384 AA.
 AC W83354;
 DT 10-FEB-1999 (first entry)
 DE Vernonia galamenensis fatty acid epoxidising enzyme.
 KW Vernonia galamenensis; fatty acid desaturase; epoxidising enzyme;
 KW expression; chimeric gene; recombinant enzyme.
 OS Vernonia galamenensis.
 PN US5846784-A.
 PD 08-DEC-1998.
 PF 11-JUN-1997; 872302.
 PR 11-JUN-1997; US-872302.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 PI Hitz WD;
 DR WPI: 99-059065/05.
 DR N-PSDB: V72550.
 PT DNA encoding Vernonia galamenensis fatty acid desaturase and fatty
 PT acid epoxidising enzyme - used to alter levels of expression of the
 PT enzymes in transformed host cells or to produce recombinant enzymes
 PS Claim 5; Column 29-32; 21pp; English.
 CC The present sequence is Vernonia galamenensis fatty acid epoxidising
 CC enzyme. The present invention also describes: (i) Vernonia galamenensis
 CC fatty acid desaturase; (ii) chimeric genes comprising the fragments
 CC linked to regulatory sequences; and (iii) transformed host cells
 CC containing the chimeric genes. The DNA's from the present invention can

CC be used to alter levels of expression of the enzymes in transformed host
 CC cells or to produce the recombinant enzymes by transformation of
 CC microbial host cells.
 SO Sequence 384 AA;

Query Match 33.2%; Score 61; DB 1; Length 384;
 Best Local Similarity 54.5%; Pred. No. 7.28e+01;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 44 CFORSAIRSSC 54
 QY 208 CWQGEIRDAC 218

RESULT 9

ID R94900 standard; Protein; 778 AA.
 AC R94900;
 DT 03-JUL-1996 (first entry)
 DE Human tatin.
 KW Trophinin; trophinin-assisting protein; tatin; bystin; lastin;
 KW embryo implantation; infertility; cell adhesion; therapy; diagnosis;
 KW tumour suppressor; cancer.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT region 516..650
 FT /note= "cystine-rich 4-tandem repeat region"
 FT modified_site 2
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 26
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 98
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 129
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 135
 FT /note= "casein kinase II phosphorylation site"
 FT modified_site 155
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 161
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 166
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 177
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 179
 FT /note= "MAP kinase phosphorylation site"
 FT modified_site 191
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 234
 FT /note= "casein kinase II phosphorylation site"
 FT modified_site 236
 FT /note= "CAMP/CGMP-dependent phosphorylation site"
 FT modified_site 237
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 269
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 274
 FT /note= "casein kinase II phosphorylation site"
 FT modified_site 296
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 350
 FT /note= "casein kinase II phosphorylation site"
 FT modified_site 363
 FT /label= CAMP/CGMP-dependent phosphorylation site
 FT modified_site 379
 FT /note= "MAP kinase phosphorylation site"
 FT modified_site 408
 FT /note= "casein kinase II phosphorylation site"
 FT modified_site 421
 FT /note= "casein kinase II phosphorylation site"
 FT modified_site 424
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 426
 FT /note= "casein kinase II phosphorylation site"

FT modified_site /note= "casein kinase II phosphorylation site"
 FT 452
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 478
 FT /note= "casein kinase II phosphorylation site"
 FT modified_site 569
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 570
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 602
 FT /note= "casein kinase II phosphorylation site"
 FT modified_site 603
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 691
 FT /note= "casein kinase II phosphorylation site"
 FT modified_site 714
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 737
 FT /note= "protein kinase C phosphorylation site"
 FT modified_site 738
 FT /note= "casein kinase II phosphorylation site"
 FT modified_site 756
 FT /note= "protein kinase C phosphorylation site"
 FT /note= "casein kinase II phosphorylation site"

WO9610414-A1.

11-APR-1996.

04-OCT-1995; U13259.

04-OCT-1994; US-317522.

12-MAY-1995; US-439818.

(LJOL-) LA JOLLA CANCER RES FOUND.

Fukuda MN;

WPI; 96-209192/21.

N-PSDB; T18677.

Mammalian trophinin and trophinin-assisting protein - used in
 inhibiting or enhancing embryo implantation, diagnosis of
 infertility and treatment of cancer

Claim 24; Fig 6; 10pp; English.

Human trophinin-assisting proteins tatin (R94900), bystin (R94902)
 and lastin (R94903) are the products of 3 cDNA clones (T18677-79,
 respectively). Expression of these proteins is required by a cell
 to effect trophinin-mediated cell adhesion. The trophinin-assisting
 proteins can function to segregate trophinin molecules (see also
 R94895) into clusters on the apical plasma membrane. They can be
 used to improve trophinin-mediated cell adhesion, e.g. to minimise
 embryo implantation failure. Tatin is probably a cytoplasmic
 protein and shows no overall significant homology to reported
 protein sequences.

Sequence 778 AA;

Query Match 33.2%; Score 61; DB 1; Length 778;

Best Local Similarity 41.2%; Pred. No. 7.28e+01;

Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 371 WLRGVSPQSCSEDPALP 387

QY 209 WQGEIRDACAEETPPP 225

RESULT 10

ID R14256 standard; Protein; 771 AA.
 AC R14256;
 DT 08-JAN-1992 (first entry)
 DE Human GCSF receptor encoded by PHQ2.
 KW Granulocyte colony stimulating factor.
 OS Homo Sapiens.
 FH Key Location/Qualifiers
 FT peptide 1..23
 FT /label= signal sequence
 FT modified_site 51..53
 FT /label= N-glycosylation
 FT modified_site 93..95
 FT /label= N-glycosylation
 FT modified_site 128..130
 FT /label= N-glycosylation

modified_site 134..136
 /label= N-glycosylation
 modified_site 389..391
 /label= N-glycosylation
 modified_site 474..476
 /label= N-glycosylation
 modified_site 570..572
 /label= N-glycosylation
 modified_site 579..581
 /label= N-glycosylation
 modified_site 712..714
 /label= N-glycosylation
 WO9114776-A.
 03-OCT-1991. J00375.
 22-MAR-1991; JP-074539.
 23-MAR-1990; JP-176629.
 03-JUL-1990; JP-176629.
 (OSAB-) OSAKA BIOSCIENCE IN.
 Nagata S, Fukunaga R;
 WPI: 91-310576/42.
 N-PSDB: Q14257.
 DNA encoding granulocyte colony stimulating factor receptor - for recombinant prodn. of GCSF receptor useful in therapy and research.
 Claim 1: Fig 8: 99pp; Japanese.
 The sequence was deduced from a DNA sequence obt'd. from a cDNA library prep'd. from human histocytic lymphoma U937 cells using DNA from the corresponding murine gene (see Q13855). The N-terminal from residue 600 - 771 is not the same as that of R14255 due to the deletion of a segment of DNA from the gene. The genes can be used to produce recombinant receptors for use in research and for diagnostic assays.
 See also R14254-R14257.
 Sequence 771 AA;

Query Match 32.6%; Score 60; DB 1; Length 771;
 Best Local Similarity 38.1%; Pred. No. 9.11e-01;
 Matches 8; Conservative 5; Mismatches 6; Indels 2; Gaps 2;

Db 669 CWRMK-RSRCGSPITAGR 688
 QY 208 CWRGEIRDACAETP-TPPKP 227

RESULT 11
 ID W20982 standard; Protein; 231 AA.
 AC W20982;
 DT 21-JUL-1997 (first entry)
 DE H. pylori secreted or periplasmic protein, hp3ell024orf6.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacterium; life cycle; activator;
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
 KW diagnosis.
 PS Helicobacter pylori.
 PN WO9640893-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996; U09122.
 PR 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaard BL;
 DR WPI: 97-052306/05.
 DR N-PSDB: T68235.
 PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) - useful for vaccines to treat or prevent H. pylori infection, and to detect Helicobacter
 PS Claim 72: Page 1373; 1481pp; English.
 CC This sequence represents a H. pylori secreted or periplasmic protein. The protein may be used in a vaccine to prevent or treat H. pylori infection or to identify H. pylori polypeptide binding compounds.
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from overlapping contigs generated by mechanically shearing the bacterial

CC DNA. The sequences were analysed for ORF of at least 180 nucleotides, and the predicted coding regions defined by computer evaluation. To identify likely H. pylori antigens for vaccine development, the amino acid sequences predicted from various ORF were analysed for significant homology to other known or exported membrane proteins. Having identified CC and determined the sequences of interest, particular regions can be CC isolated from H. pylori by PCR amplification for recombinant polypeptide CC production, e.g. in E. coli hosts.
 SQ Sequence 231 AA;

Query Match 32.1%; Score 59; DB 1; Length 231;
 Best Local Similarity 40.0%; Pred. No. 1.14e+02;
 Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Db 73 RGHSDPSKNNPGAPKPTLA 92
 QY 211 RGEIRDACAETPTPKPKLS 230

RESULT 12

ID R94957 standard; Protein; 1104 AA.

AC R94957;

DT 17-AUG-1996 (first entry)

DE NF-X1 DNA-binding protein.

KW NF-X1; transcription factor; major histocompatibility complex; MHC;

KW allergy; DNA-binding protein; diagnosis.

OS Homo sapiens.

PS Location/Qualifiers

FT domain 430..900

/label= Cysteine-rich_domain

FT region 417..468

/label= Cysteine-rich_repeat

FT region 473..521

/label= Cysteine-rich_repeat

FT region 533..583

/label= Cysteine-rich_repeat

FT region 599..647

/label= Cysteine-rich_repeat

FT region 688..737

/label= Cysteine-rich_repeat

FT region 799..839

/label= Cysteine-rich_repeat

FT region 840..890

/label= Cysteine-rich_repeat

PN WO9612823-A1.

PD 02-MAY-1996.

PF 20-OCT-1995; U12749.

PR 21-OCT-1994; US-327832.

PA (HARD) HARVARD COLLEGE.

PA (UYJO) UNIV JOHNS HOPKINS.

PI Ono SJ, Strominger JL;

DR WPI: 96-230621/23.

DR N-PSDB: T18813.

PT Transcription factor, NF-X1 and DNA encoding it - used in regulation of MHC class II expression and in treatment of allergic disease

PS Claim 6: Page 55-58; 93pp; English.

CC Human NF-X1 (R94957) is a new DNA-binding protein which regulates expression of major histocompatibility complex (MHC) class II molecules. The sequence was deduced from a cDNA clone (T18813) isolated from Raji cells. NF-X1 is a cysteine-rich polypeptide which interacts specifically with the conserved X1 box regulatory element (T18812) found in the proximal promoters of class II MHC genes. It contains a cysteine-rich domain (see also R94955-56) required for both sequence-specific binding and effector function. Overexpression of NF-X1 specifically represses transcription of the HLA-DRA gene. NF-X1 can be obt'd. by expression in transformed host cells and is useful in the treatment of allergic disease.

CC the HLA-DRA gene.

CC host cells and is useful in the treatment of allergic disease.

CC Sequence 1104 AA;

SQ

Query Match 32.1%; Score 59; DB 1; Length 1104;

Best Local Similarity 50.0%; Pred. No. 1.14e+02;

Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 415 WSRNEIPHSCGE 426
| : | : | : | : |
Qy 209 WQGEIRDACAE 220

RESULT 13
ID P50075 standard; Protein; 36 AA.
AC P50075.
DT 16-OCT-1991 (first entry)
DE Peptide with alpha-amylase-inhibitor activity.
KW Alpha-amylase-inhibitor.
OS Streptomyces aureofaciens.
FH Key Location/Qualifiers
FT disulfide_bond 9..25
FN EP-151246-A.
PD 14-AUG-1985.
PR 22-NOV-1984; 114064.
PR 21-JAN-1984; DE-402021.
PA (FARH) HOECHST AG.
PI Vertesy L, Tripiet D, Ritzel H;
DR WPI; 85-197969/33.
PT New cyclic peptide(s) - useful as alpha-amylase inhibitors.
PS Claim 3; Page 18; 20pp; German.
CC The new peptide (described as "cyclic") is useful as an alpha-amylase inhibitor, and may be used in the treatment of diabetes, prediabetes and obesity, and for preventing dental caries.
SQ Sequence 36 AA;

Query Match 31.5%; Score 58; DB 1; Length 36;
Best Local Similarity 26.7%; Pred. No. 1.42e+02;
Matches 4; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

Db 16 WRYTDVNGSCDAVT 30
| : | : | : | : |
Qy 209 WQGEIRDACAEPT 223

RESULT 14
ID W18580 standard; Protein; 919 AA.
AC W18580;
DT 19-SEP-1997 (first entry)
DE Potato alpha-glucosidase.
KW Alpha-glucosidase; potato; starch, flavour; transgenic plant.
OS Solanum tuberosum cv. Record.
PN W09724448-A1.
PD 10-JUL-1997.
PF 24-DEC-1996; G03239.
PR 28-DEC-1995; GB-026613.
PA (NICK-) NICKERSON BIOCHEM LTD.
PI Davies HV, Taylor MA;
DR WPI; 97-363683/33.
DR N-PSDB; T69925.
PT Sense and antisense plant alpha-glucosidase nucleic acids - useful for modifying alpha-glucosidase activity, e.g. for enhancing plant tissue flavour, producing novel starch, to control pre-sprouting, etc.
PS Example 1; Fig 2; 40pp; English.
CC A 105.4 kDa alpha-glucosidase (W18580) is encoded by an isolated nucleic acid (T69925) of potato cv. Record. This isolated nucleic acid can be used to modify alpha-glucosidase activity in plants or transformed microorganisms. Activity is enhanced through the use of sense nucleic acids, thereby providing more efficient breakdown of starch (useful in malting and brewing), novel starches, or enhanced flavour to plant tissues such as fruit. Inhibition of alpha-glucosidase activity through the use of antisense sequences can inhibit starch breakdown and to control pre-sprouting and dry weight/starch loss in cooking, storage and processing.
SQ Sequence 919 AA;

Query Match 31.5%; Score 58; DB 1; Length 919;
Best Local Similarity 42.9%; Pred. No. 1.42e+02;
Matches 6; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

Db 23 WKKEFRN-CDQTP 35
| : | : | : | : |
Qy 209 WQGEIRDACAEPT 222

RESULT 15
ID W87504 standard; Protein; 1061 AA.
AC W87504;
DT 23-FEB-1999 (first entry)
DE Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA24.
KW Human; N-methyl-D-aspartate receptor; NMDAR2C;
KW NMDA-activated cation-selective ion channel; glutamate receptor.
OS Homo sapiens.
PN US5849895-A.
PD 15-DEC-1998.
PF 20-APR-1994; 231193.
PR 20-APR-1994; US-231193.
PR 20-APR-1993; US-052449.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Daggett LP, Lu C;
DR WPI; 99-069812/06.
DR N-PSDB; V82910.
PT DNA encoding N-methyl-D-aspartate receptor subunit - useful for the assembly of functional glutamate receptor subunits
PS Example 3; Columns 261-268; 203pp; English.
CC The present sequence represents a human N-methyl-D-aspartate (NMDA) receptor subunit (NMDAR). The nucleic acid sequence does not contain the 860 5'-most nucleotides. Has an additional 11 nucleotides (V82891) between nucleotides 1300 and 1301, an additional 24 nucleotides (V82890) inserted between nucleotides 23050 and 2351, as set forth in V82889. The cDNA sequence is derived from clone NMDA24. The NMDAR subunits contribute to the formation of NMDA-activated cation-selective ion channels. In addition to being useful for the production of NMDA receptor subunit proteins, the nucleic acids are also useful as probes to identify and isolate nucleic acids encoding related receptor subunits. Functional glutamate receptors can be assembled from several NMDA receptor subunit proteins of one type (homomeric) or from combinations of subunit proteins of different types (heteromeric). The present invention also comprises methods for using such receptor subunits to identify and characterise compounds which affect the function of such receptors, e.g. agonists, antagonists and modulators of glutamate receptor function. The invention also comprises methods for determining whether unknown protein(s) are functional as NMDA receptor subunits.
SQ Sequence 1061 AA;

Query Match 31.5%; Score 58; DB 1; Length 1061;
Best Local Similarity 36.8%; Pred. No. 1.42e+02;
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Db 300 WRRASVWLAAMAPSPRRP 318
| : | : | : | : | : | : |
Qy 209 WQGEIRDACAEPTTPPKP 227

Search completed: Wed May 10 13:40:38 2000
Job time : 8 secs.

W O S R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 13:40:55 2000; MasPar time 47.03 Seconds
Tabular output not generated. 6.612 Million cell updates/sec

Title: >US-09-376-430-2
Description: (208-231) from US09376430A.ppt (19 of 25)
Perfect score: 184
Sequence: 1 CWMQGEIRDACAETPTPPKPKLSK 24

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 20.577; Variance 68.967; scale 0.298

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	67	36.4	610	2	US-08-865-Sequence 6, Applicatio	1.14e+01
2	66	35.9	472	1	US-08-749-Sequence 4, Applicatio	1.43e+01
3	66	35.9	472	1	US-08-749-Sequence 5, Applicatio	1.43e+01
4	64	34.8	501	2	US-09-030-Sequence 9, Applicatio	2.24e+01
5	64	34.8	501	2	US-08-467-Sequence 9, Applicatio	2.24e+01
6	64	34.8	501	1	US-08-722-Sequence 14, Applicati	2.24e+01
7	61	33.2	384	2	US-08-872-Sequence 4, Applicatio	4.35e+01
8	61	33.2	732	1	US-08-317-Sequence 5, Applicatio	4.35e+01
9	61	33.2	778	2	US-08-738-Sequence 5, Applicatio	4.35e+01
10	61	33.2	778	2	US-08-728-Sequence 5, Applicatio	4.35e+01
11	61	33.2	778	1	US-08-439-Sequence 5, Applicatio	4.35e+01
12	61	33.2	778	2	US-08-751-Sequence 5, Applicatio	4.35e+01
13	61	33.2	896	2	US-08-640-Sequence 12, Applicati	4.35e+01
14	60	32.6	771	1	US-07-923-Sequence 6, Applicatio	5.41e+01
15	59	32.1	1104	2	US-08-828-Sequence 5, Applicatio	6.72e+01
16	59	32.1	1104	2	US-08-327-Sequence 5, Applicatio	6.72e+01
17	58	31.5	54	2	US-08-660-Sequence 18, Applicati	8.35e+01
18	58	31.5	2213	1	US-08-727-Sequence 3, Applicatio	8.35e+01
19	57	31.0	472	1	US-08-749-Sequence 3, Applicatio	1.04e+02
20	57	31.0	472	1	US-08-749-Sequence 1, Applicatio	1.04e+02
21	57	31.0	1045	1	US-08-452-Sequence 2, Applicatio	1.04e+02
22	57	31.0	1052	2	US-08-852-Sequence 2, Applicatio	1.04e+02
23	57	31.0	2214	1	US-08-727-Sequence 7, Applicatio	1.04e+02

24	56	30.4	121	2	US-08-822-Sequence 4, Applicatio	1.28e+02
25	56	30.4	183	4	5168049-3 Patent No. 5168049.	1.28e+02
26	56	30.4	527	2	US-08-535-Sequence 3, Applicatio	1.28e+02
27	55	29.9	53	2	US-08-726-Sequence 144, Applicati	1.59e+02
28	55	29.9	541	3	PCT-US9550 Sequence 2, Applicatio	1.59e+02
29	55	29.9	871	2	US-08-775-Sequence 34, Applicati	1.59e+02
30	55	29.9	1940	2	US-08-644-Sequence 30, Applicati	1.59e+02
31	54	29.3	253	2	US-09-144-Sequence 10, Applicati	1.96e+02
32	54	29.3	253	2	US-08-685-Sequence 10, Applicati	1.96e+02
33	54	29.3	280	2	US-08-464-Sequence 6, Applicatio	1.96e+02
34	54	29.3	292	2	US-08-246-Sequence 6, Applicatio	1.96e+02
35	54	29.3	450	2	US-08-396-Sequence 2, Applicatio	1.96e+02
36	54	29.3	559	2	US-08-884-Sequence 6, Applicatio	1.96e+02
37	54	29.3	622	2	US-08-459-Sequence 2, Applicatio	1.96e+02
38	54	29.3	699	2	US-08-800-Sequence 7, Applicatio	1.96e+02
39	54	29.3	947	2	US-09-023-Sequence 2, Applicatio	1.96e+02
40	54	29.3	987	2	US-08-436-Sequence 6, Applicatio	1.96e+02
41	54	29.3	1276	1	US-08-222-Sequence 24, Applicati	1.96e+02
42	54	29.3	2843	1	US-07-741-Sequence 2, Applicatio	1.96e+02
43	54	29.3	2843	1	US-08-289-Sequence 2, Applicatio	1.96e+02
44	54	29.3	2843	1	US-08-452-Sequence 2, Applicatio	1.96e+02
45	54	29.3	2973	2	US-09-003-Sequence 7, Applicatio	1.96e+02

ALIGNMENTS

RESULT 1
ID US-08-865-337A-6 STANDARD; PRT; 610 AA.
XX
AC xxxxxx
DT

Sequence 6, Application US/08865337A

Sequence 6, Application US/08865337A
Patent No. 5972649
GENERAL INFORMATION:

APPLICANT: Au-Young, Janice
APPLICANT: Covitz, Peter
APPLICANT: Tang, Y. Tom
TITLE OF INVENTION: HUMAN MULTIPLE ENDOCRINE NEOPLASIA TYPE
TITLE OF INVENTION: 1 PROTEIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/865,337A
FILING DATE: Herewith
CLASSIFICATION: 53b
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0305 US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 6:

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 610 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC LIBRARY: GenBank
CC CLONE: 1945387
SQ SEQUENCE 610 AA; 67466 MW; 1878918 CN;

Query Match 36.4%; Score 67; DB 2; Length 610;
Best Local Similarity 45.8%; Pred. No. 1.14e+01;
Matches 11; Conservative 5; Mismatches 7; Indels 1; Gaps 1;

Db 479 RRGPRRESKPEPPPPKPKALDK 502
QY 209 WQGEIRDACAETPTPPK-PKLSK 231

RESULT 2
ID US-08-749-903-4 STANDARD; PRT; 472 AA.

XXXXXX

Sequence 4, Application US/08749903

Sequence 4, Application US/08749903
Patent No. 5759812

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Hawkins, Phillip R.

TITLE OF INVENTION: NOVEL HUMAN SELENIUM-BINDING PROTEIN

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/749,903

FILING DATE: Filed Herewith

PRIOR APPLICATION NUMBER:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0163 US

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 472 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 227630

SEQUENCE 472 AA; 52289 MW; 1178083 CN;

Query Match

35.9%; Score 66; DB 1; Length 472;

Best Local Similarity 36.4%; Pred. No. 1.43e+01;
Matches 8; Conservative 7; Mismatches 5; Indels 2; Gaps 2;

Db 327 WLHGDIQ-Y-DISNPQKPLA 346

QY 209 WQGEIRDACAETPTPPKPKLS 230

RESULT 3
ID US-08-749-903-5 STANDARD; PRT; 472 AA.

XXXXXX

Sequence 5, Application US/08749903

Sequence 5, Application US/08749903
Patent No. 5759812

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Hawkins, Phillip R.

TITLE OF INVENTION: NOVEL HUMAN SELENIUM-BINDING PROTEIN

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/749,903

FILING DATE: Filed Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0163 US

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 472 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 298710

SEQUENCE 472 AA; 52408 MW; 1200217 CN;

Query Match 35.9%; Score 66; DB 1; Length 472;
Best Local Similarity 36.4%; Pred. No. 1.43e+01;
Matches 8; Conservative 7; Mismatches 5; Indels 2; Gaps 2;

Db 327 WLHGDIQ-Y-DISNPQKPLA 346

QY 209 WQGEIRDACAETPTPPKPKLS 230

RESULT 4

ID US-09-030-582-9 STANDARD; PRT; 501 AA.

AC	xxxxxx
XX	
DT	
DE	
XX	
Sequence 9, Application US/09030582	
XX	
Sequence 9, Application US/09030582	
CC	Patent No. 5994506
CC	GENERAL INFORMATION:
CC	APPLICANT: SOPPET, DANIEL R
CC	TITLE OF INVENTION: ADRENERGIC RECEPTOR
CC	NUMBER OF SEQUENCES: 13
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
CC	ADDRESSES: Stewart & Olstein
CC	STREET: 6 Becker Farm Road
CC	CITY: Roseland
CC	STATE: NJ
CC	COUNTRY: USA
CC	ZIP: 07068-1739
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Floppy disk
CC	COMPUTER: IBM PC compatible
CC	OPERATING SYSTEM: PC-DOS/MS-DOS
CC	SOFTWARE: PatentIn Release #1.0, Version #1.30
CC	CURRENT APPLICATION DATA:
CC	FILING DATE: 08/09/030,582
CC	PRIOR APPLICATION NUMBER: US/09/030,582
CC	CLASSIFICATION:
CC	PRIOR APPLICATION DATA: 08/467,568
CC	APPLICATION NUMBER: 08/467,568
CC	FILING DATE:
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: Ferraro, Gregory D
CC	REGISTRATION NUMBER: 36,134
CC	REFERENCE/DOCKET NUMBER: 325800-324
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: 201-994-1700
CC	TELEFAX: 201-994-1744
CC	INFORMATION FOR SEQ ID NO: 9:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 501 amino acids
CC	TYPE: amino acid
CC	STRANDEDNESS: single
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: protein
CC	SEQUENCE 501 AA; 54681 MW; 1272839 CN;
SQ	
Query Match	34.8%; Score 64; DB 2; Length 501;
Best Local Similarity	31.8%; Pred. No. 2.24e+01;
Matches	7; Conservative 6; Mismatches 8; Indels 1;
Dd	457 CWGRSGDPRPSCAPKSPACRTR 478
QY	208 CWQR-GEIRDACAETPTPPKPK 228
RESULT	5
ID	US-08-467-568-9 STANDARD; PRT; 501 AA.
XX	xxxxxx
DT	
DE	
XX	
Sequence 9, Application US/08467568	
CC	Patent No. 5817477
CC	GENERAL INFORMATION:
CC	APPLICANT: SOPPET, DANIEL R
CC	TITLE OF INVENTION: ADRENERGIC RECEPTOR
CC	NUMBER OF SEQUENCES: 13
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: Merck & Co., Inc.
CC	STREET: 126 Lincoln Avenue
CC	CITY: Rahway
CC	STATE: New Jersey
CC	COUNTRY: United States of America
CC	ZIP: 07065
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Floppy disk
CC	COMPUTER: IBM PC compatible
CC	OPERATING SYSTEM: PC-DOS/MS-DOS
CC	SOFTWARE: PatentIn Release #1.0, Version #1.25
CC	CURRENT APPLICATION DATA:
CC	FILING DATE: 06-JUN-1995
CC	CLASSIFICATION: 514
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: Ferraro, Gregory D
CC	REGISTRATION NUMBER: 36,134
CC	REFERENCE/DOCKET NUMBER: 325800-324
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: 201-994-1700
CC	TELEFAX: 201-994-1744
CC	INFORMATION FOR SEQ ID NO: 9:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 501 amino acids
CC	TYPE: amino acid
CC	STRANDEDNESS: single
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: protein
CC	SEQUENCE 501 AA; 54681 MW; 1272839 CN;
SQ	
Query Match	34.8%; Score 64; DB 2; Length 501;
Best Local Similarity	31.8%; Pred. No. 2.24e+01;
Matches	7; Conservative 6; Mismatches 8; Indels 1; Gaps 1;
Dd	457 CWGRSGDPRPSCAPKSPACRTR 478
QY	208 CWQR-GEIRDACAETPTPPKPK 228
RESULT	5
ID	US-08-467-568-9 STANDARD; PRT; 501 AA.
XX	xxxxxx
DT	
DE	
XX	
Sequence 9, Application US/08467568	
CC	Patent No. 5817477
CC	GENERAL INFORMATION:
CC	APPLICANT: SOPPET, DANIEL R
CC	TITLE OF INVENTION: ADRENERGIC RECEPTOR
CC	NUMBER OF SEQUENCES: 13
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: Merck & Co., Inc.
CC	STREET: 126 Lincoln Avenue
CC	CITY: Rahway
CC	STATE: New Jersey
CC	COUNTRY: United States of America
CC	ZIP: 07065
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Floppy disk
CC	COMPUTER: IBM PC compatible
CC	OPERATING SYSTEM: PC-DOS/MS-DOS
CC	SOFTWARE: PatentIn Release #1.0, Version #1.25
CC	CURRENT APPLICATION DATA:
CC	FILING DATE: 06-JUN-1995
CC	CLASSIFICATION: 514
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: Ferraro, Gregory D
CC	REGISTRATION NUMBER: 36,134
CC	REFERENCE/DOCKET NUMBER: 325800-324
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: 201-994-1700
CC	TELEFAX: 201-994-1744
CC	INFORMATION FOR SEQ ID NO: 9:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 501 amino acids
CC	TYPE: amino acid
CC	STRANDEDNESS: single
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: protein
CC	SEQUENCE 501 AA; 54681 MW; 1272839 CN;
SQ	
Query Match	34.8%; Score 64; DB 2; Length 501;
Best Local Similarity	31.8%; Pred. No. 2.24e+01;
Matches	7; Conservative 6; Mismatches 8; Indels 1; Gaps 1;
Dd	457 CWGRSGDPRPSCAPKSPACRTR 478
QY	208 CWQR-GEIRDACAETPTPPKPK 228
RESULT	5
ID	US-08-467-568-9 STANDARD; PRT; 501 AA.
XX	xxxxxx
DT	
DE	
XX	
Sequence 9, Application US/08467568	
CC	Patent No. 5817477
CC	GENERAL INFORMATION:
CC	APPLICANT: SOPPET, DANIEL R
CC	TITLE OF INVENTION: ADRENERGIC RECEPTOR
CC	NUMBER OF SEQUENCES: 13
CC	CORRESPONDENCE ADDRESS:
CC	

CC APPLICATION NUMBER: US/08/722,001
CC FILING DATE:
CC CLASSIFICATION: 514
CC PRIOR APPLICATION NUMBER: 08/229,276
CC FILING DATE: 14-APR-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Appollina, Mary A.
CC REGISTRATION NUMBER: 34,087
CC REFERENCE/DOCKET NUMBER: 19169Y
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (908)594-3462
CC TELEFAX: (908)594-4720
CC TELEX: 138825
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 501 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 501 AA; 54681 MW; 1272839 CN;

Query Match 34.8%; Score 64; DB 1; Length 501;
Best Local Similarity 31.8%; Pred. No. 2.24e+01;
Matches 7; Conservative 6; Mismatches 8; Indels 1; Gaps 1;

Db 457 CWGRSGDPRPCAPKSPACRTR 478
QY 208 CQWR-GEIRDCAETPTPPKPK 228

RESULT 7
ID US-08-872-302-4 STANDARD; PRT; 384 AA.
XX xxxxxx
AC
DT
DT
XX
DE
XX
CC Sequence 4, Application US/08872302
CC Sequencé 4, Application US/08872302
CC Patent No. 5846784
CC GENERAL INFORMATION:
CC APPLICANT: Hitz, William D
CC TITLE OF INVENTION: Fatty Acid Modifying Enzymes From
CC TITLE OF INVENTION: Developing Seeds of Vernonia galamensis
CC NUMBER OF SEQUENCES: 10
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: E.I. duPont de Nemours and Co.
CC STREET: 1007 Market Street
CC CITY: Wilmington
CC STATE: Delaware
CC COUNTRY: USA
CC ZIP: 19898
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Majarian, William R
CC REGISTRATION NUMBER: P-41,173
CC REFERENCE/DOCKET NUMBER: BB-1084
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 302-992-4926

CC APPLICATION NUMBER: US/08/722,001
CC FILING DATE:
CC CLASSIFICATION: 514
CC PRIOR APPLICATION NUMBER: 08/229,276
CC FILING DATE: 14-APR-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Appollina, Mary A.
CC REGISTRATION NUMBER: 34,087
CC REFERENCE/DOCKET NUMBER: 19169Y
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (908)594-3462
CC TELEFAX: (908)594-4720
CC TELEX: 138825
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 501 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 501 AA; 54681 MW; 1272839 CN;

Query Match 34.8%; Score 64; DB 1; Length 501;
Best Local Similarity 31.8%; Pred. No. 2.24e+01;
Matches 7; Conservative 6; Mismatches 8; Indels 1; Gaps 1;

Db 457 CWGRSGDPRPCAPKSPACRTR 478
QY 208 CQWR-GEIRDCAETPTPPKPK 228

RESULT 8
ID US-08-317-522A-5 STANDARD; PRT; 732 AA.
XX xxxxxx
AC
DT
DT
XX
DE
XX
CC Sequence 5, Application US/08317522A
CC Sequencé 5, Application US/08317522A
CC Patent No. 5599918
CC GENERAL INFORMATION:
CC APPLICANT: Fukuda, Michiko N.
CC TITLE OF INVENTION: Trophinin and Trophinin-Assisting
CC TITLE OF INVENTION: Proteins
CC NUMBER OF SEQUENCES: #13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Campbell and Flores
CC STREET: 4370 La Jolla Village Drive, Suite 700
CC CITY: San Diego
CC STATE: California
CC COUNTRY: USA
CC ZIP: 92122
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/317,522A
CC FILING DATE: 04-OCT-1994
CC CLASSIFICATION: 536
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Campbell, Cathryn A.
CC REGISTRATION NUMBER: 31,815
CC REFERENCE/DOCKET NUMBER: P-LA 9991
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 535-9001
CC TELEFAX: (619) 535-8949
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 732 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 732 AA; 78879 MW; 2817009 CN;

Query Match 33.2%; Score 61; DB 1; Length 732;
Best Local Similarity 41.2%; Pred. No. 4.35e+01;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 371 WLKRGVSQSCSDPALP 387
QY 209 WQGEIRDCAETPTPP 225

```

RESULT 9
ID US-08-738-975-5 STANDARD; PRT; 778 AA.
XX AC
XX
XX
DT
XX
XX
DE
XX
Sequence 5, Application US/08738975
Patent No. 5880267
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/738.975
FILING DATE: herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/439,818
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/317,522
FILING DATE: 04-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 2251
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 778 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 778 AA; 83758 MW; 3141054 CN;

Query Match 33.2%; Score 61; DB 2; Length 778;
Best Local Similarity 41.2%; Pred. No. 4.35e+01;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 371 WLRGVSPQSCSDPALP 387
| | | : | | | : |
QY 209 WQGEIRDACAETPTTP 225

RESULT 10
ID US-08-728-626-5 STANDARD; PRT; 778 AA.
XX AC
XX
XX
DT
XX
XX
DE
XX
Sequence 5, Application US/08728626
Patent No. 5910451
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/738.975
FILING DATE: herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/439,818
FILING DATE: 05-Dec-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 2251
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 778 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 778 AA; 83758 MW; 3141054 CN;

Query Match 33.2%; Score 61; DB 2; Length 778;
Best Local Similarity 41.2%; Pred. No. 4.35e+01;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 371 WLRGVSPQSCSDPALP 387
| | | : | | | : |
QY 209 WQGEIRDACAETPTTP 225

RESULT 11
ID US-08-439-818A-5 STANDARD; PRT; 778 AA.
XX AC
XX
XX
DT
XX
XX
DE
XX
Sequence 5, Application US/08439818A
Patent No. 5654145
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA: P-LA 1563
CC APPLICATION NUMBER: US/08/439,818A
CC FILING DATE: 12-MAY-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/317,522
CC FILING DATE: 04-OCT-1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Campbell, Cathryn A.
CC REGISTRATION NUMBER: 31,815
CC REFERENCE/DOCKET NUMBER: P-LA 1563
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 535-9001
CC TELEFAX: (619) 535-8949
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 778 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 778 AA; 83758 MW; 3141054 CN;
Query Match 33.2%; Score 61; DB 1; Length 778;
Best Local Similarity 41.2%; Pred. No. 4.35e+01;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
Db 371 WLRGVSPQSCSEDPALP 387
QY 209 WQGEIRDACAEPTTPP 225
||| :||:| :|
RESULT 12
ID US-08-751-965-5 STANDARD; PRT; 778 AA.
XX
AC xxxxxx
DT
DE
XX
Sequence 5, Application US/08751965
Sequence 5, Application US/08751965
Patent No. 5858360
GENERAL INFORMATION:
CC APPLICANT: Fukuda, Michiko N.
CC TITLE OF INVENTION: Trophinin and Trophinin-Assisting
CC TITLE OF INVENTION: Proteins
CC NUMBER OF SEQUENCES: 22
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Campbell and Flores
CC STREET: 4370 La Jolla Village Drive, Suite 700
CC CITY: San Diego
CC STATE: California
CC COUNTRY: USA
CC ZIP: 92122
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/751,965
CC FILING DATE: Herewith
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/439,818
CC FILING DATE: 12-MAY-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Campbell, Cathryn A.
CC REGISTRATION NUMBER: 31,815
CC REFERENCE/DOCKET NUMBER: P-LA 2252

CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 535-9001
CC TELEFAX: (619) 535-8949
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 778 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 778 AA; 83758 MW; 3141054 CN;
Query Match 33.2%; Score 61; DB 2; Length 778;
Best Local Similarity 41.2%; Pred. No. 4.35e+01;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
Db 371 WLRGVSPQSCSEDPALP 387
QY 209 WQGEIRDACAEPTTPP 225
||| :||:| :|
RESULT 13
ID US-08-640-389A-12 STANDARD; PRT; 896 AA.
XX
AC xxxxxx
DT
DE
XX
Sequence 12, Application US/08640389A
Sequence 12, Application US/08640389A
Patent No. 5912123
GENERAL INFORMATION:
CC APPLICANT: Snodgrass, H. R.
CC APPLICANT: Cioffi, Joseph
CC APPLICANT: Zupancic, Thomas J.
CC APPLICANT: Shafer, Alan W.
CC TITLE OF INVENTION: DETECTION OF THE LEPTIN
CC TITLE OF INVENTION: RECEPTOR IN REPRODUCTIVE ORGANS AND METHODS FOR
CC TITLE OF INVENTION: REGULATING REPRODUCTIVE BIOLOGY
CC NUMBER OF SEQUENCES: 16
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Pennie & Edmonds LLP
CC STREET: 1155 Avenue of the Americas
CC CITY: New York
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 10036-2711
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/640,389A
CC FILING DATE: 29-APR-1996
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Poissant, Brian M.
CC REGISTRATION NUMBER: 28,462
CC REFERENCE/DOCKET NUMBER: 8907-032
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 790-9090
CC TELEFAX: (212) 869-9741/8864
CC TELEX: 66141 PENNIE
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 896 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC SEQUENCE 896 AA; 101628 MW; 4594631 CN;

WISORH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 13:40:03 2000; MasPar time 4.83 Seconds
Tabular output not generated. 234.590 Million cell updates/sec

Title: >US-09-376-430-2
Description: (208-231) from US09376430A.pap (19 of 25)
Perfect Score: 184
Sequence: 1 CQWGEIRDACAETPTPPKPLSK 24

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 28.631; Variance 43.457; scale 0.659

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description	Pred. No.
1	71	38.6	768	2	DNA ligase (ATP) (EC 6.5.1.1)	1.73e+01
2	70	38.0	1433	1	M polyprotein precurs	2.61e-01
3	69	37.5	79	2	hypothetical protein	3.93e-01
4	69	37.5	334	1	E2 protein - bovine p	3.93e-01
5	67	36.4	233	2	hypothetical protein	8.84e-01
6	67	36.4	1557	2	protein CNK - fruit f	8.84e-01
7	66	35.9	472	2	selenium-binding prot	1.32e+00
8	66	35.9	876	2	serine/threonine-spec	1.32e+00
9	65	35.3	505	2	hypothetical protein	1.96e+00
10	64	34.8	501	2	alpha-1A-adrenergic r	2.90e+00
11	64	34.8	1288	2	mitogen-activated pro	2.90e+00
12	63	34.2	106	2	hypothetical protein	4.27e+00
13	63	34.2	352	1	adaptive response reg	4.27e+00
14	63	34.2	392	1	serine proteinase eas	4.27e+00
15	63	34.2	408	1	gene 16 protein - hum	4.27e+00
16	63	34.2	1017	1	replication licensing	4.27e+00
17	62	33.7	83	1	posterior pituitary p	6.27e+00
18	62	33.7	48	1	probable Mead involve	6.27e+00
19	62	33.7	101	2	hypothetical protein	6.27e+00
20	62	33.7	179	2	hypothetical protein	6.27e+00
21	62	33.7	191	2	hypothetical protein	6.27e+00
22	62	33.7	203	2	hypothetical protein	6.27e+00
23	62	33.7	252	2	ORF 5' of calmodulin	6.27e+00

24	62	33.7	340	2	S32584	structural protein VP	6.27e+00
25	62	33.7	344	2	H75099	hypothetical protein	6.27e+00
26	62	33.7	1209	2	T00373	hypothetical protein	6.27e+00
27	62	33.7	2020	2	C48399	ABC-type transport pr	6.27e+00
28	61	33.2	246	2	I50127	MHC class II histocom	9.16e+00
29	61	33.2	304	2	T02125	chlorophyll a/b-bindl	9.16e+00
30	61	33.2	444	2	T01548	protein kinase homolo	9.16e+00
31	61	33.2	778	2	I38487	tastin - human	9.16e+00
32	61	33.2	1091	2	E71322	isoleucine--tRNA liga	9.16e+00
33	60	32.6	197	2	F65019	hypothetical protein	1.33e+01
34	60	32.6	230	2	A36509	deoxyribonuclease I (1.33e+01
35	60	32.6	235	2	A38966	deoxyribonuclease I (1.33e+01
36	60	32.6	303	1	PHUNE	Na+/K+-exchanging ATP	1.33e+01
37	60	32.6	367	2	I40843	heat shock protein dn	1.33e+01
38	60	32.6	391	2	T08338	hypothetical protein	1.33e+01
39	60	32.6	524	2	S47143	A mating type protein	1.33e+01
40	60	32.6	771	2	B38252	granulocyte colony-st	1.33e+01
41	60	32.6	856	1	A45394	env polyprotein precu	1.33e+01
42	60	32.6	2907	2	A57278	fibrillin-2 precursor	1.33e+01
43	59	32.1	286	2	A70854	hypothetical protein	1.93e+01
44	59	32.1	790	2	S50082	nuclear cap binding p	1.93e+01
45	59	32.1	1104	2	I38869	NFX1 - human	1.93e+01

ALIGNMENTS

RESULT 1
ENTRY A29066 #type complete
TITLE DNA ligase (ATP) (EC 6.5.1.1) - fission yeast
ALTERNATE_NAMES (Schizosaccharomyces pombe)
ORGANISM polydeoxyribonucleotide synthase (ATP)
#formal_name Schizosaccharomyces pombe
DATE 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 16-Jul-1999
ACCESSIONS A29066
REFERENCE A29066
#authors Barker, D.G.; White, J.H.M.; Johnston, L.H.
#journal Eur. J. Biochem. (1987) 162:659-667
#title Molecular characterization of the DNA ligase gene, CDC17, from the fission yeast Schizosaccharomyces pombe.
#cross-references MUID:87161837
#accession A29066
#molecule_type DNA
#residues 1-768 #label BAR
#cross-references GB:X05107; NID:g4924; PIDN:CAA28754.1; PID:g4925

GENETICS 28/2; 197/3
#introns
CLASSIFICATION #superfamily yeast polydeoxyribonucleotide synthase (ATP)
KEYWORDS ligase; nucleus; phosphoprotein
FEATURE 416
#active_site Lys (covalent AMP-binding) #status predicted
SUMMARY #length 768 #molecular-weight 86580 #checksum 7185
Query Match 38.6%; Score 71; DB 2; Length 768;
Best Local Similarity 47.6%; Pred. No. 1.73e-01;
Matches 10; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db 374 GTLRETCKLTGPIPKPLAK 394
Qy 212 GEIRDACAETPTPPKPLSK 231
RESULT 2
ENTRY GNVUBW #type complete
TITLE M polyprotein precursor - bunyamwera virus
CONTAINS .glycoprotein G1; glycoprotein G2; nonstructural protein
ORGANISM #formal_name bunyamwera virus
DATE 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 16-Jul-1999
ACCESSIONS A04101
REFERENCE A04101
#authors Less, J.F.; Pringle, C.R.; Elliott, R.M.

```

#journal      Virology (1986) 148:1-14
#title        Nucleotide sequence of the Bunyamvera virus M RNA segment:
               conservation of structural features in the bunyavirus
               glycoprotein gene product.
#cross-references MUID:86098655
#accession    A04101
               #molecule_type genomic RNA
               #residues 1-1433 #label LEE
               ##cross-references GB:M1852; NID:g210743; PIDN:AAA42777.1; PID:g210744
               ##residues 1-1433 #label LEE
               #COMMENT Specific enzymatic cleavages in vivo yield mature proteins
               including nonstructural protein NS-M, glycoprotein G1, and
               glycoprotein G2. However, exact cleavage sites are undetermined.
GENETICS
#map_position segment M
CLASSIFICATION #superfamily bunyavirus M polypotein
KEYWORDS       glycoprotein; nonstructural protein; polypotein;
               transmembrane protein
FEATURE
1-16          #domain signal sequence #status predicted #label SIG\
17-1433       #product M polypotein #status predicted #label MPP\
60,248,624,1169 #binding-site carbohydrate (Asn) (covalent) #status
               predicted
SUMMARY        #length 1433 #molecular-weight 162077 #checksum 1114
Query Match    38.0%; Score 70; DB 1; Length 1433;
Best Local Similarity 31.6%; Pred. No. 2,61e-01;
Matches 6; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Db 481 CWKNEELKEDCVGLIAPK 499
||| |::: | ||
QY 208 CWRGEIRDCAETPTPK 226

RESULT 3
ENTRY #type complete
TITLE hypothetical protein 5 - Streptococcus phage phi-O1205
ORGANISM #formal_name Streptococcus phage phi-O1205
DATE 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change
24-Sep-1999
ACCESSIONS T13293
REFERENCE #authors Stanley, E.; Fitzgerald, G.F.; Le Marrec, C.; Fayard, B.; van
Sinderen, D.
#journal Microbiology (1997) 143:3417-3429
#title Sequence analysis and characterization of phi O1205, a
temperate bacteriophage infecting Streptococcus
thermophilus CNRZ1205.
#cross-references MUID:98048466
#accession T13293
#status translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-79 #label STA
#cross-references EMBL:U88974; NID:g2444080; PID:g2444085;
#experimental_source host Streptococcus thermophilus strain CNRZ1205
#length 79 #molecular-weight 9335 #checksum 4997
SUMMARY
Query Match 37.5%; Score 69; DB 2; Length 79;
Best Local Similarity 44.4%; Pred. No. 3.93e-01;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Db 44 CWOQSEISKACELLQIPS 61
|||:::| |
QY 208 CWRGEIRDCAETPTPK 225

RESULT 4
ENTRY #type complete
TITLE E2 protein - bovine papillomavirus type 4
ORGANISM #formal_name bovine papillomavirus type 4
DATE 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change
16-Jul-1999
ACCESSIONS F27129

```

```

REFERENCE A92795
#authors Patel, K.R.; Smith, K.T.; Campo, M.S.
#journal J. Gen. Virol. (1987) 68:2117-2128
#title The nucleotide sequence and genome organization of bovine
papillomavirus type 4.
#cross-references MUID:87282264
#accession F27129
               #molecule_type DNA
               #residues 1-334 #label PAT
               ##cross-references GB:D00146; NID:g222360; PIDN:BAA00098.1;
               ##residues 1-334 #label PAT
               ##cross-references GB:D00146; NID:g222360; PID:g222365
CLASSIFICATION #superfamily papillomavirus E2 protein
KEYWORDS       DNA binding; early protein; transcription regulation
SUMMARY        #length 334 #molecular-weight 37203 #checksum 5374
Query Match    37.5%; Score 69; DB 1; Length 334;
Best Local Similarity 50.0%; Pred. No. 3.93e-01;
Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

Db 205 RRRGTPRERCPGTPPTTP 224
||| |::: | ||
QY 209 WQRGE-IRDCAETPTPK 227

RESULT 5
ENTRY #type complete
TITLE hypothetical protein Rv1100 - Mycobacterium tuberculosis
(h37rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
20-Sep-1999
ACCESSIONS B70897
REFERENCE #authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eigimeier, K.; Gas, S.; Barry,
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, L.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession B70897
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-233 #label COL
#cross-references GB:AL021897; GB:AL123456; NID:g3256022; PID:e1251979;
PID:g2896737
#experimental_source strain H37Rv
GENETICS
#gene Rv1100
CLASSIFICATION #superfamily Mycobacterium tuberculosis hypothetical protein
Rv1100
SUMMARY        #length 233 #molecular-weight 24563 #checksum 8118
Query Match    36.4%; Score 67; DB 2; Length 233;
Best Local Similarity 38.1%; Pred. No. 8.84e-01;
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Db 15 WDTGHVTAEPQTPPAKPL 35
||| |::: | ||
QY 209 WQRGE-IRDCAETPTPK 229

RESULT 6
ENTRY #type complete
TITLE protein CNK - fruit fly (Drosophila melanogaster)

```



```

ALTERNATE_NAMES   ksr connector enhancer protein
ORGANISM           #formal_name Drosophila melanogaster
DATE               13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change
ACCESSIONS        T13160
REFERENCE          217616
#authors           Therrien, M.; Wong, A.M.; Rubin, G.M.
#journal           Cell (1998) 95:343-353
#title            CNK, a RAF-binding multidomain protein required for RAS
                  signaling.
#cross-references  MUID:99029842
#accession         T13160
#status            preliminary; translated from GB/EMBL/DBJ
#molecule_type    mRNA
#residues          1-1557 #label THE
#cross-references  EMBL:AF100152; NID:g3930778; PID:g3930779;
                  PIDN:AAC80557.1
FUNCTION           regulates RAF function and/or targets RAF to a specific
                  subcellular compartment upon RAS activation
SUMMARY            #length 1557 #molecular-weight 171446 #checksum 5203
Query Match       36.4%; Score 67; DB 2; Length 1557;
Best Local Similarity 50.0%; Pred. No. 8.84e-01;
Matches           7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
Db 371 SCSDPTPTDPKLA 384
      ||| |||
QY 217 ACATPTTPPKPLS 230

RESULT 7
ENTRY   S27878 #type complete
TITLE   selenium-binding protein, hepatic - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE     17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
ACCESSIONS S27878
REFERENCE   S27878
#authors    Bansel, M.P.; Cook, R.; Oborn, C.J.; Scott, J.; Mukhopadhyay,
            T.; Medina, D.
#submission submitted to the EMBL Data Library, February 1991
#description Molecular cloning and sequencing of a liver protein which
            binds selenium.
#accession  S27878
#molecule_type mRNA
#residues   1-472 #label BAN
#cross-references EMBL:M32032; NID:g200951; PID:g200952
REFERENCE   JN0271
#authors    Pumford, N.R.; Martin, B.M.; Hinson, J.A.
#journal     Biochem. Biophys. Res. Commun. (1992) 182:1348-1355
#title       A metabolite of acetaminophen covalently binds to the 56 kDa
            selenium binding protein.
#cross-references MUID:92171951
#accession  JN0271
#molecule_type protein
#residues   72-73,'K';175-181,'G',183,'S',185-189;196-220;228-242;
            290-296;334-343;399-407,'D' #label PUM
COMMENT     This protein covalently binds a metabolite of acetaminophen. This
            binding correlates with the hepatotoxicity.
KEYWORDS    liver; selenium binding
SUMMARY      #length 472 #molecular-weight 52351 #checksum 1422
Query Match     35.9%; Score 66; DB 2; Length 472;
Best Local Similarity 36.4%; Pred. No. 1.32e-00;
Matches         8; Conservative 7; Mismatches 5; Indels 2; Gaps 2;
Db 327 WLHGDIQ-Y-DISNPQKPLA 346
      ||| |||
QY 209 WQGEIRDACAETPTTPPKPLS 230

RESULT 8

```

```

ENTRY   T00540 #type complete
TITLE   serine/threonine-specific protein kinase (EC 2.7.1.-)
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
          cress
DATE     01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change
ACCESSIONS T00540
REFERENCE   214167
#authors    Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby,
            M.L.; Brandon, R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage,
            A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
#submission submitted to the EMBL Data Library, July 1997
#description Arabidopsis thaliana chromosome II BAC T20K24 genomic
            sequence.
#accession  T00540
#status     translated from GB/EMBL/DBJ
#molecule_type DNA
#residues   1-876 #label ROU
#cross-references EMBL:AC002392; NID:g3176701; PIDN:AAD12037.1;
                  PID:g3176723
#experimental_source cultivar Columbia
GENETICS
#map_position 29/1; 211/2; 372/1; 419/2; 443/2; 467/2; 494/2; 548/1; 617/3;
#introns      660/1; 681/1; 744/2
#note         T20K24.21
CLASSIFICATION #superfamily unassigned Ser/Thr or Tyr-specific protein
               kinases; protein kinase homology
KEYWORDS       phosphotransferase
FEATURE        574-849
SUMMARY        #length 876 #molecular-weight 98711 #checksum 37
Query Match    35.9%; Score 66; DB 2; Length 876;
Best Local Similarity 36.4%; Pred. No. 1.32e+00;
Matches        8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;
Db 815 WKMEIYALACTESIAQRTMS 836
      ||| |||
QY 209 WQGEIRDACAETPTTPPKPLS 230

RESULT 9
ENTRY   S64837 #type complete
TITLE   hypothetical protein YLR015w - yeast (Saccharomyces
          cerevisiae)
ALTERNATE_NAMES hypothetical protein LI583
ORGANISM         #formal_name Saccharomyces cerevisiae
DATE             01-Aug-1995 #sequence_revision 24-May-1996 #text_change
ACCESSIONS       S64837; S59274; S48527
REFERENCE        S64742
#authors         Vandenbol, M.; Portetelle, D.; Hilger, F.
#submission      submitted to the Protein Sequence Database, May 1996
#accession       S64837
#molecule_type DNA
#residues        1-505 #label VAN
#cross-references EMBL:Z73187; NID:gl360313; PID:e245494; PID:gl360314;
                  MIPS:YLR015w
#experimental_source strain S288C
REFERENCE        S59270
#authors         Saville, S.P.; Atkinson, S.; Jamieson, L.; Pocklington, M.J.;
            Orr, E.
#submission      submitted to the EMBL Data Library, August 1995
#description     A 7.8kb fragment from chromosome XII of Saccharomyces
            cerevisiae does not harbour PKC2.
#accession       S59274
#molecule_type DNA
#residues        1-464,'ISLL',468 #label SAV
#cross-references EMBL:X90564; NID:g975221; PID:g975226
#experimental_source strain S288C
REFERENCE        S48527

```

```

#authors      Levin, D.E.; Stevenson, W.D.
#submission   submitted to the EMBL Data Library, July 1994
#description   The S.cerevisiae PKC2 does not exist in the yeast genome.
#accession    S48527
#molecule_type DNA
#residues     1-472 ##label LEV
#cross-references EMBL:L34405

GENETICS
#map_position 12R
SUMMARY      #length 505 #molecular-weight 58347 #checksum 3214
Query Match 35.3%; Score 65; DB 2; Length 505;
Best Local Similarity 42.9%; Pred. No. 1.9e+00;
Matches 9; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

Db 392 WQHGARDNDKNTSAKKK 412
||| ||| : ||| |
QY 209 WQGEIRDACAT-TPPKPK 228

RESULT 10
ENTRY   #type complete
TITLE   alpha-1A-adrenergic receptor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE     31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
ACCESSIONS
REFERENCE JH0447
#authors  Bruno, J.F.; Whittaker, J.; Song, J.; Berelowitz, M.
#journal  Biochem. Biophys. Res. Commun. (1991) 179:1485-1490
#title    Molecular cloning and sequencing of a cDNA encoding a human
          alpha 1A adrenergic receptor.
#cross-references MUID:92028892
#accession  JH0447
#molecule_type mRNA
#residues  1-501 ##label BRU
#cross-references GB:M76446; NID:g177806; PIDN:AAA35496.1; PID:g177807
#experimental_source hippocampus
CLASSIFICATION #superfamily vertebrate rhodopsin
KEYWORDS       G protein-coupled receptor; transmembrane protein
FEATURE
56-79          #domain transmembrane #status predicted #label TM1\
93-116         #domain transmembrane #status predicted #label TM2\
128-151        #domain transmembrane #status predicted #label TM3\
172-196        #domain transmembrane #status predicted #label TM4\
210-233        #domain transmembrane #status predicted #label TM5\
308-331        #domain transmembrane #status predicted #label TM6\
339-363        #domain transmembrane #status predicted #label TM7\
SUMMARY      #length 501 #molecular-weight 54681 #checksum 3876

Query Match 34.8%; Score 64; DB 2; Length 501;
Best Local Similarity 31.8%; Pred. No. 2.9e+00;
Matches 7; Conservative 6; Mismatches 8; Indels 1; Gaps 1;

Db 457 CWGRSGDRPSCAPKSPACRTR 478
||| ||| : ||| : :
QY 208 CWQR-GEIRDACAT-TPPKPK 228

RESULT 11
ENTRY   #type complete
TITLE   mitogen-activated protein kinase kinase kinase (EC 2.7.-.-) -
          human
ORGANISM #formal_name Homo sapiens #common_name man
DATE     23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change
ACCESSIONS
REFERENCE JE0363
#authors  Wang, X.; Diener, K.; Tan, T.; Yao, Z.
#journal  Biochem. Biophys. Res. Commun. (1998) 253:33-37
#title    MAPKKK6, a novel mitogen-activated protein kinase kinase
          kinase, that associates with MAPKKK5.
#accession JE0363

```

```

#status      preliminary
#molecule_type mRNA
#residues    1-1288 ##label WAN
#cross-references GB:AF100318
KEYWORDS     phosphotransferase
SUMMARY      #length 1288 #molecular-weight 142502 #checksum 4392

Query Match 34.8%; Score 64; DB 2; Length 1288;
Best Local Similarity 52.9%; Pred. No. 2.9e+00;
Matches 9; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Db 75 LREACAQVPRPPPOL 91
||| ||| : ||| |
QY 214 IRDACAETPT-PPKPKL 229

RESULT 12
ENTRY   #type complete
TITLE   hypothetical protein APE1518 - Aeropyrum pernix (strain K1)
ORGANISM #formal_name Aeropyrum pernix
DATE     20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
ACCESSIONS
REFERENCE G72632
#authors  Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.;
          Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba,
          S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai,
          Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
          Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
          Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura,
          Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
#journal  DNA Res. (1999) 6:83-101
#title    Complete genome sequence of an aerobic hyper-thermophilic
          Crenarchaeon, Aeropyrum pernix K1.
#cross-references MUID:99310339
#accession  G72632
#status      preliminary
#molecule_type DNA
#residues    1-106 ##label KAW
#cross-references DBJ:AP000061; NID:g5104821; PIDN:BAA80517.1;
          PID:d1044303; PID:g5105203
#experimental_source strain K1

GENETICS
#gene       APE1518
SUMMARY     #length 106 #molecular-weight 11245 #checksum 923
Query Match 34.2%; Score 63; DB 2; Length 106;
Best Local Similarity 26.1%; Pred. No. 4.27e+00;
Matches 6; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

Db 46 CSSRSVARGVAPVAPSSPRRLP 68
||| ||| : ||| : |
QY 208 CWQGEIRDACAT-TPPKPKLS 230

RESULT 13
ENTRY   #type complete
TITLE   adaptive response regulatory protein - Salmonella typhimurium
          methylated-DNA--protein-cysteine S-methyltransferase (EC
          2.1.1.63) ada; methylphosphotriester-DNA methyltransferase
          #formal_name Salmonella typhimurium
          31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change
          13-Aug-1999
ACCESSIONS
REFERENCE A39433
#authors  Hakura, A.; Mörimoto, K.; Sofuni, T.; Nohmi, T.
          J. Bacteriol. (1991) 173:3663-3672
#title    Cloning and characterization of the salmonella typhimurium
          ada gene, which encodes O(6)-methylguanine-DNA
          methyltransferase.
#cross-references MUID:91267628
#accession  A39433
#molecule_type DNA

```

```

##residues      1-352 #label HAK
##cross-references GB:D90221; NID:g217046; PIDN:BAA14252.1;

COMMENT  This enzyme is part of an inducible DNA repair system that protects
          against methylating and alkylating agents by transferring the
          alkyl group from O(6)-alkylguanine in DNA to a cysteine residue
          located within the enzyme itself.

GENETICS
#gene
CLASSIFICATION #superfamily adaptive response regulatory protein;
                methylated-DNA--protein-cysteine S-methyltransferase
                homology; methylphosphotriester-DNA methyltransferase
                homology
KEYWORDS  DNA binding; DNA repair; methylated amino acid;
                methyltransferase; transcription regulation
FEATURE   6-190
#domain methylphosphotriester-DNA methyltransferase
                homology #label MPTN
267-347  #domain methylated-DNA--protein-cysteine
                S-methyltransferase homology #label MGTN
320      #binding_site methyl (Cys) (covalent) #status predicted
SUMMARY   #length 352 #molecular-weight 39217 #checksum 4195

Query Match      34.2%; Score 63; DB 1; Length 352;
Best Local Similarity 53.8%; Pred. No. 4.27e+00;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 11 CMLRVOARDASD 23
Qy 208 CQORGEIRDACAE 220

RESULT 14
ENTRY serine proteinase easter (EC 3.4.21.-) precursor - fruit fly
TITLE (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
10-Sep-1999
ACCESSIONS A30100; A32727
REFERENCE A30100
#authors Chasan, R.; Anderson, K.V.
#journal Cell (1989) 56:391-400
#title The role of Easter, an apparent serine protease, in
        organizing the dorsal-ventral pattern of the Drosophila
        embryo.
##cross-references MUID:89119561
#accession A30100
##molecule_type mRNA
##residues 1-392 #label CHA
##cross-references GB:J03154; NID:g157313; PIDN:AAA28496.1; PID:g157314
REFERENCE A32727
#authors Jin, Y.; Anderson, K.V.
#journal Cell (1990) 60:873-881
#title Dominant and recessive alleles of the Drosophila easter gene
        are point mutations at conserved sites in the serine
        protease catalytic domain.
##cross-references MUID:90182675
#accession A32727
##status preliminary; nucleic acid sequence not shown; not
        compared with conceptual translation
##molecule_type DNA
##residues 128-392 #label JIN

GENETICS
#gene FlyBase:ea
##cross-references FlyBase:FBgn0000533
CLASSIFICATION #superfamily serine proteinase easter; trypsin homology
KEYWORDS hydrolase; serine proteinase
FEATURE 1-19
          #domain signal sequence #status predicted #label SIGN
20-392  #domain easter protein #status predicted #label MATN
128-386 #domain trypsin homology #label TRY
173,240,338 #active_site His, Asp, Ser #status predicted

```

```

SUMMARY      #length 392 #molecular-weight 43065 #checksum 8517

Query Match      34.2%; Score 63; DB 1; Length 392;
Best Local Similarity 43.8%; Pred. No. 4.27e+00;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 94 RESSETTPPPKPNVT 109
Qy 215 RDACAETPTPPKPKLS 230

RESULT 15
ENTRY W2BE16 #type complete
TITLE gene 16 protein - human herpesvirus 3
ORGANISM #formal_name human herpesvirus 3, varicella-zoster virus
DATE 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change
16-Jul-1999
ACCESSIONS G27342
REFERENCE A27345
#authors Davison, A.J.; Scott, J.E.
#journal J. Gen. Virol. (1986) 67:1759-1816
#title The complete DNA sequence of varicella-zoster virus.
#cross-references MUID:86306657
#accession G27342
##molecule_type DNA
##residues 1-408 #label DAV
##cross-references EMBL:X04370; NID:g59989; PIDN:CAA27899.1; PID:g60005

GENETICS
#gene 16
CLASSIFICATION #superfamily varicella-zoster virus gene 16 protein
KEYWORDS DNA binding
SUMMARY #length 408 #molecular-weight 46090 #checksum 765

Query Match      34.2%; Score 63; DB 1; Length 408;
Best Local Similarity 40.9%; Pred. No. 4.27e+00;
Matches 9; Conservative 1; Mismatches 11; Indels 1; Gaps 1;

Db 371 WQEGIGIREYCVAPPVDPAGTL 392
Qy 209 WQRG-EIRDACAETPTPPKPKL 229

Search completed: Wed May 10 13:40:12 2000
Job time : 9 secs.

```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 13:33:40 2000; MasPar time 87.63 Seconds
Tabular output not generated. 8.341 Million cell updates/sec

Title: >US-09-376-430-2
Description: (208-231) from US09376430A.pap (19 of 25)
Perfect Score: 184
Sequence: 1 CWORGEIRDACAETPTPKPKLSK 24

Scoring table:
PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 29.362; Variance 39.672; scale 0.740

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	71	38.6	768	1	DNL1_SCHPO	DNA LIGASE (EC 6.5.1.1)
2	70	38.0	1433	1	VGLM_BUNYW	M POLYPROTEIN PRECURSOR
3	69	37.5	334	1	VE2_BPVA	REGULATORY PROTEIN E2
4	67	36.4	611	1	MEN1_MOUSE	MENIN
5	67	36.4	615	1	MEN1_HUMAN	MENIN
6	66	35.9	472	1	SBP2_MOUSE	SELENIUM-BINDING PROTE
7	66	35.9	472	1	SBP1_MOUSE	SELENIUM-BINDING PROTE
8	65	35.3	505	1	YLIR_YEAST	HYPOTHETICAL 58.3 KD P
9	63	34.2	352	1	ADA_SALTY	ADA REGULATORY PROTEIN
10	63	34.2	392	1	EAST_DROME	SERINE PROTEASE EASTER
11	63	34.2	408	1	VPAP_VZVD	DNA POLYMERASE PROCESS
12	63	34.2	1017	1	MCW6_YEAST	MINICHROMOSOME MAINTEN
13	62	33.7	48	1	POPI_BOVIN	POSTERIOR PITUITARY PE
14	62	33.7	179	1	VG04_HSVII	HYPOTHETICAL GENE 4 PR
15	62	33.7	252	1	YCAL_PLAFA	HYPOTHETICAL PROTEIN I
16	62	33.7	340	1	VP35_EBOV	POLYMERASE COMPLEX PRO
17	61	33.2	446	1	NH19_CAEEL	STEROID HORMONE RECEPT
18	61	33.2	517	1	MTCL_COREQ	MODIFICATION METHYLASE
19	61	33.2	778	1	TAST_HUMAN	TASTIN (TROPHININ-ASSI
20	61	33.2	1091	1	SYL_TREPA	ISOEUCYL-TRNA SYNTHET
21	60	32.6	210	1	DRNG_AERHY	EXTRACELLULAR DEOXYRIB
22	60	32.6	230	1	DRNE_VIBCH	EXTRACELLULAR DEOXYRIB
23	60	32.6	235	1	END1_ECOLI	ENDONUCLEASE I PRECURS

24	60	32.6	303	1	ATNB_HUMAN	SODIUM/POTASSIUM-TRANS	5.01e+00
25	60	32.6	367	1	DNAJ_COXBU	DNAJ PROTEIN	5.01e+00
26	60	32.6	856	1	ENV_FIVU1	ENV POLYPROTEIN PRECUR	5.01e+00
27	60	32.6	2907	1	FBN2_MOUSE	FIBRILLIN 2 PRECURSOR	5.01e+00
28	59	32.1	157	1	YSAA_ECOLI	PUTATIVE ELECTRON TRAN	7.51e+00
29	59	32.1	338	1	CITC_LEUMC	[CITRATE (PRO-35)-LYAS	7.51e+00
30	59	32.1	790	1	CB80_HUMAN	80 KD NUCLEAR CAP BIND	7.51e+00
31	59	32.1	892	1	SYA_METJA	ALANYL-TRNA SYNTHETASE	7.51e+00
32	59	32.1	1104	1	NFX1_HUMAN	TRANSCRIPTIONAL REPRES	7.51e+00
33	59	32.1	2911	1	FBN2_HUMAN	FIBRILLIN 2 PRECURSOR	7.51e+00
34	58	31.5	36	1	IAA_STRAU	ALPHA-AMYLASE INHIBITO	1.12e+01
35	58	31.5	371	1	LIG2_PCHACH	LIGNINASE LG2 PRECURSO	1.12e+01
36	58	31.5	423	1	GRF_MOUSE	GROWTH HORMONE-RELEASI	1.12e+01
37	58	31.5	440	1	NAM1_YEAST	NAM1 PROTEIN PRECURSOR	1.12e+01
38	58	31.5	453	1	CA44_BOVIN	COLLAGEN ALPHA 4(IV) C	1.12e+01
39	58	31.5	455	1	GIPR_RAT	GASTRIC INHIBITORY POL	1.12e+01
40	58	31.5	462	1	GIPR_MESAU	GASTRIC INHIBITORY POL	1.12e+01
41	58	31.5	466	1	GIPR_HUMAN	GASTRIC INHIBITORY POL	1.12e+01
42	58	31.5	881	1	Y310_HUMAN	HYPOPHETICAL PROTEIN K	1.12e+01
43	58	31.5	1247	1	NIDO_HUMAN	NIDOGEN PRECURSOR (ENT	1.12e+01
44	58	31.5	1437	1	VGLM_BUNGE	M POLYPROTEIN PRECURSO	1.12e+01
45	57	31.0	302	1	PHEB_PESP	CATECHOL 1,2-DIOXYGENA	1.66e+01

ALIGNMENTS

RESULT 1 STANDARD: PRT; 768 AA.
AC P12000;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DNA LIGASE (EC 6.5.1.1) (POLYDEOXYRIBONUCLEOTIDE SYNTHASE [ATP]).
GN CDC17 OR SPAC57A10.12C OR SPAC2068.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87161837.
RA Barker D.G., White J.H.M., Johnston L.H.;
RT "Molecular characterisation of the DNA ligase gene, CDC17, from the
fission yeast Schizosaccharomyces pombe.";
RL Eur. J. Biochem. 162:659-667(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA Badcock K., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN SEALS DURING DNA REPLICATION, DNA
RECOMBINATION AND DNA REPAIR NICKS IN DOUBLE-STRANDED DNA.
CC -1- CATALYTIC ACTIVITY: ATP + (DEOXYRIBONUCLEOTIDE)(N) +
(DEOXYRIBONUCLEOTIDE)(M) = AMP + PYROPHOSPHATE +
(DEOXYRIBONUCLEOTIDE)(N+M).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT DNA LIGASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05107; CAA28754.1; -
DR EMBL; Z94864; CAB08176.1; -
DR EMBL; Z95334; CAB08595.1; -
DR FIC; A29066; A29066.
DR PROSITE; PS00333; DNA_LIGASE_A2; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
DR PFAM; PF01068; DNA_ligase; 1.
KW DNA repair; DNA replication; DNA recombination; Cell division; Ligase;

KW ATP-binding; Nuclear protein. AMP (BY SIMILARITY).
FT BINDING 416 416
SQ SEQUENCE 768 AA; 86581 MW; 6783FF3DDC675F31 CRC64;

Query Match 38.6%; Score 71; DB 1; Length 768;
Best Local Similarity 47.6%; Pred. No. 4.27e-02;
Matches 10; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db 374 GTLRETKLTPGTPKPLAK 394
| : : : | | | : : |
QY 212 GEIRDACAETP-TPPKKLSK 231

RESULT 2
ID VGLM_BUNYV STANDARD; PRT; 1433 AA.
AC P04505;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE M POLYPROTEIN PRECURSOR [CONTAINS: NONSTRUCTURAL PROTEIN NS-M;
DE GLYCOPROTEINS G1 AND G2].
GN M.
OS Bunyamwera virus.
OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Bunyavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86098655.
RA Lees J.F., Pringle C.R., Elliott R.M.;
RT "Nucleotide sequence of the Bunyamwera virus M RNA segment:
RT conservation of structural features in the Bunyavirus glycoprotein
RT gene product";
RL Virology 148:1-14(1986).
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS
CC INCLUDING NONSTRUCTURAL PROTEIN NS-M, GLYCOPROTEIN G1, AND
CC GLYCOPROTEIN G2.
CC -!- SIMILARITY: BELONGS TO THE BUNYAVIRUSES M POLYPROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M1852; AAA42777.1; -
DR PIR; A04101; GNVUBW.
KW Polyprotein; Glycoprotein; Transmembrane; Nonstructural protein;
KW Signal.
FT SIGNAL 1 16
FT CHAIN 17 1433 M POLYPROTEIN.
FT CHAIN 17 302 GLYCOPROTEIN G2.
FT CHAIN 303 477 NONSTRUCTURAL PROTEIN NS-M.
FT CHAIN 478 1433 GLYCOPROTEIN G1.
FT TRANSMEM 1382 1409 POTENTIAL.
FT CARBOHYD 60 60 POTENTIAL.
FT CARBOHYD 248 248 POTENTIAL.
FT CARBOHYD 624 624 POTENTIAL.
FT CARBOHYD 1169 1169 POTENTIAL.
SQ SEQUENCE 1433 AA; 162077 MW; CDG1ABDE782018E0 CRC64;

Query Match 38.0%; Score 70; DB 1; Length 1433;
Best Local Similarity 31.6%; Pred. No. 6.73e-02;
Matches 6; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Db 481 CWNKEELKEDCVGLIAPK 499
| : : : | | : : |
QY 208 CWORGEIRDAETP-TPPK 226

RESULT 3
ID VE2_BPV4
AC P08345;
STANDARD; PRT; 334 AA.

DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE REGULATORY PROTEIN E2.
GN E2.
OS Bovine papillomavirus type 4.
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87282264.
RA Patel K.R., Smith K.T., Campo M.S.;
RT "The nucleotide sequence and genome organization of bovine
RT papillomavirus type 4";
RL J. Gen. Virol. 68:2117-2128(1987).
CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCCNNNNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -!- SUBUNIT: BINDS DNA AS A DIMER.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X05817; -; NOT_ANNOTATED_CDS.
DR EMBL; D00146; BAA00038.1; -
DR PIR; F27129; W2WL84.
DR HSSP; F03122; 2BOP.
DR TRANSFAC; T00205; -
DR PFAM; PF00508; E2_N; 1.
DR PFAM; PF00511; E2_C; 1.
KW Early protein; Transcription regulation; Activator; DNA-binding;
KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ SEQUENCE 334 AA; 37203 MW; 061C6A317B5CDB1F CRC64;

Query Match 37.5%; Score 69; DB 1; Length 334;
Best Local Similarity 50.0%; Pred. No. 1.06e-01;
Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

Db 205 RRGTPERERCPGTPPTPP 224
: : : : | : : : : |
QY 209 WORGE-IRDAETPTPKP 227

RESULT 4
ID MEN1_MOUSE STANDARD; PRT; 611 AA.
AC O88559;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE MENIN.
GN MEN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/Ola;
RA Bassett J.H.D., Thakker R.V.;
RT "Genomic sequence of mouse MEN1 gene";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99039765.

RA Stewart C., Parente F., Pighl F., Farnebo F., Quincey D., Silins G.,
RA Bergman L., Carle G.F., Lemmens I., Grimmond S., Xian C.Z., Khodai S.,
RA Teh B.T., Lagercrantz J., Siggers P., Calender A., van de Vem V.,
RA Kas K., Weber G., Hayward N., Gaudray P., Larsson C.;
RA "Characterization of the mouse Men1 gene and its expression during
RA development.";
RA Oncogene 17:2485-2493(1998).
CC CC -1- FUNCTION: NOT KNOWN.
CC CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC CC -1- TISSUE SPECIFICITY: UBIQUITOUS. EXPRESSED AT HIGH LEVEL IN TESTIS
CC AND CNS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----
CC DR EMBL; AF072755; AAC26001.1; -;
CC DR EMBL; AF093756; AAC78843.1; -;
CC DR EMBL; AF016398; AAC79938.1; -;
CC KW Nuclear protein.
CC FT CONFLICT 457 457 M -> I (IN REF. 2).
CC FT CONFLICT 466 466 G -> E (IN REF. 2).
CC FT CONFLICT 512 512 L -> S (IN REF. 2).
CC SQ SEQUENCE 611 AA; 67472 MW; C3FBB3A5695244E9 CRC64;

Query Match 36.4%; Score 67; DB 1; Length 611;
Best Local Similarity 45.8%; Pred. No. 2.57e-01;
Matches 11; Conservative 5; Mismatches 7; Indels 1; Gaps 1;

Db 479 RRRGPRRESKPEPPPPKPKALDK 502
:|:| | :|:| | | | | | | |
QY 209 WQGEIRDACAETPTPK-PKLSK 231

RESULT 5
ID MEN1_HUMAN STANDARD; PRT; 615 AA.
AC O00255; O00632;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE MEN1.
GN MEN1 OR SCG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS FMEN1.
RC TISSUE=LEUKOCYTES;
RX MEDLINE; 97258940;
RA Chandrasekharappa S.C., Guru S.C., Manickam P., Olufemi S.-E.,
RA Collins F.S., Emmert-Buck M.R., Skarulis M.C., Debelenko L.V., Zhuang Z.,
RA Lubensky I.A., Liotta L.A., Crabtree J.S., Wang Y., Roe B.A.,
RA Weisemann J., Boguski M.S., Agarwal S.K., Kester M.B., Kim Y.S.,
RA Heppner C., Dong Q., Spiegel A.M., Burns A.L., Marx S.J.;
RA "Positional cloning of the gene for multiple endocrine neoplasia-type
RA 1.";
RA Science 276:404-407(1997).
RL [2]
RP VARIANTS FMEN1 AND SPORADIC MEN1.
RX MEDLINE; 97358593.
RA Emmert-Buck M.R., Skarulis M.C., Doppman J.L., Kim Y.S.,
RA Lubensky I.A., Zhuang Z., Green J.S., Guru S.C., Manickam P.,
RA Olufemi S.-E., Liotta L.A., Chandrasekharappa S.C., Collins F.S.,
RA Spiegel A.M., Burns A.L., Marx S.J.;
RA "Germline mutations of the MEN1 gene in familial multiple endocrine
RA neoplasia type 1 and related states.";
RA Hum. Mol. Genet. 6:1169-1175(1997).
RN [3]

RP VARIANT PARATHYROID ADENOMA LYS-26.
RX MEDLINE; 97385243.
RA Heppner C., Kester M.B., Agarwal S.K., Debelenko L.V.,
RA Emmert-Buck M.R., Guru S.C., Manickam P., Olufemi S.-E.,
RA Skarulis M.C., Doppman J.L., Alexander R.H., Kim Y.S., Saggat S.K.,
RA Lubensky I.A., Zhuang Z., Liotta L.A., Chandrasekharappa S.C.,
RA Collins F.S., Spiegel A.M., Burns A.L., Marx S.J.;
RA "Somatic mutation of the MEN1 gene in parathyroid tumours.";
RA Nat. Genet. 16:375-378(1997).
RL [4]
RP VARIANTS FMEN1 D-42; P-165; D-169; S-188 AND E-289.
RX MEDLINE; 98130524.
RA Basset J.H.D., Forbes S.A., Pannett A.A.J., Lloyd S.E.,
RA Christie P.T., Wooding B., Harding B., Besser G.M., Edwards C.R.,
RA Monson J.P., Sampson J., Wass J.A.H., Wheeler M.H., Thakker R.V.;
RA "Characterization of mutations in patients with multiple endocrine
RA neoplasia type 1.";
RA Am. J. Hum. Genet. 62:232-244(1998).
RN [5]
RP VARIANTS FMEN1
RX MEDLINE; 98349969.
RA Giraud S., Zhang C.X., Serova-Sinilnikova O., Wautot V., Salandre J.,
RA Buissou N., Waterlot C., Bauders C., Porchet N., Aubert J.-P., Emy P.,
RA Cadiot G., Delmer B., Chabre O., Niccoli P., Leprat F., Duron F.,
RA Emperauger B., Cougard P., Goudet P., Saifati E., Riou J.-P.,
RA Guichard S., Rodier M., Meyrier A., Rohmer V., Vantghem M.-C.,
RA Assayag M., Peix J.-L., Pugeat M., Caron P., Vantghem M.-C.,
RA Gaudray P., Proye C., Conte-Devolx B., Chanson P., Shugart Y.Y.,
RA Goldgar D., Murat A., Calender A.;
RA "Germline mutation analysis in patients with multiple endocrine
RA neoplasia type 1 and related disorders.";
RA Am. J. Hum. Genet. 63:455-467(1998).
RN [6]
RP VARIANT FIHP LYS-260.
RX MEDLINE; 99011276.
RA Teh B.T., Esapa C.T., Houlston R., Grandell U., Farnebo F.,
RA Nordenskjold M., Pearce C.J., Carmichael D., Larsson C., Harris P.E.;
RA "A family with isolated hyperparathyroidism segregating a missense
RA MEN1 mutation and showing loss of the wild-type alleles in the
RA parathyroid tumors.";
RA Am. J. Hum. Genet. 63:1544-1549(1998).
RL [7]
RP VARIANT FIHP GLU-189.
RX MEDLINE; 99057176.
RA Fujimori M., Shirahama S., Sakurai A., Hashizume K., Hama Y., Ito K.,
RA Shingu K., Kobayashi S., Anano J., Fukushima Y.;
RA "Novel V184E MEN1 germline mutation in a Japanese kindred with
RA familial hyperparathyroidism.";
RA Am. J. Med. Genet. 80:221-222(1998).
RN [8]
RP VARIANTS FMEN1
RX MEDLINE; 98334342.
RA Agarwal S.K., Debelenko L.V., Kester M.B., Guru S.C., Manickam P.,
RA Olufemi S.-E., Skarulis M.C., Heppner C., Crabtree J.S.,
RA Lubensky I.A., Zhuang Z., Kim Y.S., Chandrasekharappa S.C.,
RA Collins F.S., Liotta L.A., Spiegel A.M., Burns A.L., Emmert-Buck M.R.,
RA Marx S.J.;
RA "Analysis of recurrent germline mutations in the MEN1 gene encountered
RA in apparently unrelated families.";
RA Hum. Mutat. 12:75-82(1998).
RL [9]
RP VARIANT FMEN1 ILE-135 AND LYS-364.
RX MEDLINE; 98410971.
RA Boeni R., Vortmeyer A.O., Pack S., Park W.-S., Burg G., Hofbauer G.,
RA Darling T., Liotta L., Zhuang Z.;
RA "Somatic mutations of the MEN1 tumor suppressor gene detected in
RA sporadic angiofibromas.";
RA J. Invest. Dermatol. 111:539-540(1998).
RN [10]
RP VARIANTS FMEN1 LYS-119 DEL AND GLN-171--LEU-173 DEL.
RX MEDLINE; 98419173.
RA Sakurai A., Shirahama S., Fujimori M., Katali M., Itakura Y.,
RA Kobayashi S., Anano J., Fukushima Y., Hashizume K.;

DR EMBL; S56599; CAB31472.1; -.
 DR MGD; MGI:104859; SELENBP2.
 KW Selenium.
 SQ SEQUENCE 472 AA; 52628 MW; C32FE819C4AD07CA CRC64;
 Query Match 35.9%; Score 66; DB 1; Length 472;
 Best Local Similarity 36.4%; Pred. No. 3.98e-01;
 Matches 8; Conservative 7; Mismatches 5; Indels 2; Gaps 2;
 Db 327 WLHGDIRQ-Y-DISNPQPRLT 346
 QY 209 WQGEIRDACAETPTPPPKLS 230
 RESULT 7
 ID SBPL_MOUSE STANDARD; PRT; 472 AA.
 AC P17563;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE SELENIUM-BINDING PROTEIN 1 (56 KD SELENIUM-BINDING PROTEIN) (SP56).
 GN SELENBP1 OR LPSB.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE; 91029855.
 RA Bansel M.P., Mukhopadhyay T., Scott J., Cook R.G., Mukhopadhyay R.,
 RA Medina D.;
 RT "DNA sequencing of a mouse liver protein that binds selenium:
 RT implications for selenium's mechanism of action in cancer
 RT prevention.";
 RL Carcinogenesis 11:2071-2073(1990).
 CC -!- FUNCTION: NOT KNOWN; BIND SELENIUM.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LIVER, KIDNEY AND, TO A
 CC LESSER EXTENT, LUNG.
 CC -!- SIMILARITY: BELONGS TO THE SELENIUM-BINDING PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M32032; AAA40104.1; -.
 DR PIR; S27878; S27878.
 DR SWISS-2DPAGE; P17563; MOUSE.
 DR MGD; MGI:96825; SELENBP1.
 KW Selenium.
 SQ SEQUENCE 472 AA; 52352 MW; D501292C4876033D CRC64;
 Query Match 35.9%; Score 66; DB 1; Length 472;
 Best Local Similarity 36.4%; Pred. No. 3.98e-01;
 Matches 8; Conservative 7; Mismatches 5; Indels 2; Gaps 2;
 Db 327 WLHGDIRQ-Y-DISNPQPRLA 346
 QY 209 WQGEIRDACAETPTPPPKLS 230
 RESULT 8
 ID YLIR_YEAST STANDARD; PRT; 505 AA.
 AC P43132;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL 58.3 KD PROTEIN IN PP1-SNF7 INTERGENIC REGION (ORFX).
 GN YLR015W.

OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Vandenbol M., Portetelle D., Hilger F.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-472 FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE; 93179524.
 RA Levin D.E., Stevenson W.D., Watanabe M.;
 RT "Evidence against the existence of the purported Saccharomyces
 RT cerevisiae PKC2 gene.";
 RL Curr. Biol. 4:990-995(1994).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z73187; CAA97537.1; -.
 DR EMBL; L34405; AAA34835.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 505 AA; 58347 MW; D5B33221E9F10379 CRC64;
 Query Match 35.3%; Score 65; DB 1; Length 505;
 Best Local Similarity 42.9%; Pred. No. 6.14e-01;
 Matches 9; Conservative 4; Mismatches 7; Indels 1; Gaps 1;
 Db 392 WQHGGEARDESNKNTTSAKKK 412
 QY 209 WQGEIRDACAET-PTPPKPK 228
 RESULT 9
 ID ADA_SALTY STANDARD; PRT; 352 AA.
 AC P26189;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ADA REGULATORY PROTEIN (REGULATORY PROTEIN OF ADAPTATIVE RESPONSE)
 DE [CONTAINS: METHYLATED-DNA--PROTEIN-CYSTEINE METHYLTRANSFERASE
 DE (EC 2.1.1.63) (O-6-METHYLGUANINE-DNA ALKYLTRANSFERASE)].
 GN ADA.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 91267928.
 RA Hakura A., Morimoto K., Sofuni T., Nohmi T.;
 RT "Cloning and characterization of the Salmonella typhimurium ada gene,
 RT which encodes O6-methylguanine-DNA methyltransferase.";
 RL J. Bacteriol. 173:3663-3672(1991).
 CC -!- FUNCTION: REPAIR OF ALKYLATED GUANINE IN DNA BY STOICHIOMETRICALLY
 CC TRANSFERRING THE ALKYL GROUP AT THE O-6 POSITION TO A CYSTEINE
 CC RESIDUE IN THE ENZYME. THIS IS A SUICIDE REACTION: THE ENZYME IS
 CC IRREVERSIBLY INACTIVATED. CAN ALSO REPAIR O-4-METHYLTHYMINE.
 CC -!- FUNCTION: THE METHYLATED ADA PROTEIN ACTS AS A POSITIVE REGULATOR
 CC OF ITS OWN SYNTHESIS, AS WELL AS THAT OF OTHER PROTEINS. THE
 CC TRANSCRIPTION-ACTIVATING FUNCTION OF THE ADA PROTEIN RESIDES IN
 CC ITS N-TERMINUS. IT ACTIVATES THE TRANSCRIPTION OF ALKA, ALKB AND
 CC AIDB.
 CC -!- CATALYTIC ACTIVITY: DNA (CONTAINING O6-METHYLGUANINE) + PROTEIN
 CC L-CYSTEINE = DNA (WITHOUT O6-METHYLGUANINE) + PROTEIN S-METHYL-
 CC L-CYSTEINE.
 CC -!- SIMILARITY: THE ENZYME PART SHOWS SIMILARITY TO CORRESPONDING
 CC ENZYMES IN PROKARYOTES AND EUKARYOTES.
 CC -!- SIMILARITY: THE REGULATORY PART BELONGS TO THE ARAC/XYS FAMILY OF

TRANSCRIPTIONAL REGULATORS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: D90221; BAA14252.1; -
 DR PIR: A39433; XIEBOT.
 DR HSSP: P06134; ISFE.
 DR STYGENE; SG10002; ADA.
 DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 2.
 DR PROSITE; PS00374; MGMT; 1.
 DR PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
 DR PFAM; PF00165; HTH_2; 1.
 DR PFAM; PF01035; Methyltransf_1; 1.
 DR Transcription regulation; Activator; DNA-binding; Transferase;
 KW Methyltransferase; DNA repair; Zinc; Metal-binding.
 CHAIN 179 352
 FT METAL 37 37 METHYLTRANSFERASE.
 FT METAL 41 41 BY SIMILARITY.
 FT METAL 68 68 BY SIMILARITY.
 FT METAL 71 71 BY SIMILARITY.
 FT SITE 127 128 CLEAVAGE (BY SIMILARITY).
 FT SITE 177 178 CLEAVAGE (BY SIMILARITY).
 FT ACT_SITE 68 68 ACCEPTOR FOR METHYL FROM PHOSPHOTRIESTER
 FT ACT_SITE 320 320 (BY SIMILARITY).
 FT DNA_BIND 101 120 ALKYL GROUP ACCEPTOR (BY SIMILARITY).
 FT H-T-H MOTIF (BY SIMILARITY).
 SQ SEQUENCE 352 AA; 39217 MW; B21C1BF3CBD14F20 CRC64;

Query Match 34.2%; Score 63; DB 1; Length 352;
 Best Local Similarity 53.8%; Pred. No. 1.44e+00;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 11 CQLRVQARDASD 23
 |||:||||:
 QY 208 CWQGEIRDACAE 220

RESULT 10
 ID EAST_DROME STANDARD; PRT; 392 AA.
 AC P13582;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE SERINE PROTEASE EASTER PRECURSOR (EC 3.4.21.-).
 GN EA.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 89119561.
 RA Chasan R., Anderson K.V.;
 RT "The role of easter, an apparent serine protease, in organizing the
 RT dorsal-ventral pattern of the Drosophila embryo.";
 RL Cell 56:391-400(1989).
 RN [2]
 RP MUTAGENESIS.
 RX MEDLINE; 90182675.
 RA Jin Y., Anderson K.V.;
 RT "Dominant and recessive alleles of the Drosophila easter gene are
 RT point mutations at conserved sites in the serine protease catalytic
 RT domain.";
 RL Cell 60:873-881(1990).
 CC -1- FUNCTION: EASTER IS AN EXTRACYTOSOLIC SERINE PROTEASE WHICH
 CC PLAYS A ROLE IN THE ORGANIZATION OF THE DORSAL-VENTRAL PATTERN OF

CC THE DROSOPHILA EMBRYO. THE PROTEOLYTIC CLEAVAGE CATALYZED BY
 CC EASTER MAY BE A CRITICAL, REGULATED STEP THAT DEFINES CELL
 CC IDENTITIES ALONG THE DORSAL-VENTRAL CONTINUUM.
 CC -1- MISCELLANEOUS: IT IS POSSIBLE THAT EITHER TOLL OR SNAKE ENCODES
 CC THE SUBSTRATE FOR THE EASTER PROTEASE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: J03154; AAA28496.1; -
 DR PIR: A30100; A30100.
 DR PIR: A32727; A32727.
 DR HSSP: P00763; IDPO.
 DR FLYBASE; FBgn0000533; ea.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR PFAM; PF00089; trypsin; 1.
 KW Developmental protein; Serine protease; Hydrolase; Zymogen; Signal.
 FT SIGNAL 1 19 PROBABLE.
 FT PROPEP 20 127 ACTIVATION PEPTIDE (POTENTIAL).
 FT CHAIN 128 392 SERINE PROTEASE EASTER.
 FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 240 240 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 338 338 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 158 174 BY SIMILARITY.
 FT DISULFID 307 324 BY SIMILARITY.
 FT DISULFID 334 367 BY SIMILARITY.
 FT MUTAGEN 131 131 G->E: DOMINANT ALLELE EA 20N.
 FT MUTAGEN 172 172 S->L: RECESSIVE ALLELE EA 111.
 FT MUTAGEN 283 283 G->S: DOMINANT ALLELE EA 84B.
 FT MUTAGEN 324 324 C->Y: RECESSIVE ALLELE EA 818.
 FT MUTAGEN 325 325 A->V: DOMINANT ALLELES EA 831 & EA 4102.
 FT MUTAGEN 335 335 R->C: DOMINANT ALLELE EA 125.3.
 FT MUTAGEN 336 336 G->S: DOMINANT ALLELE EA 12A.
 FT MUTAGEN 338 338 S->A: DESTROY ACTIVITY.
 FT MUTAGEN 339 339 G->R: RECESSIVE ALLELE EA 1.
 FT MUTAGEN 360 360 V->M: DOMINANT ALLELE EA 5.13.
 FT MUTAGEN 363 363 G->E: RECESSIVE ALLELE EA 8.
 FT MUTAGEN 371 371 G->R: DOMINANT ALLELE EA 161.13.
 FT MUTAGEN 373 373 P->S: DOMINANT ALLELE EA 5022.
 SQ SEQUENCE 392 AA; 43065 MW; 1309C41E6FC176B6 CRC64;

Query Match 34.2%; Score 63; DB 1; Length 392;
 Best Local Similarity 43.8%; Pred. No. 1.44e+00;
 Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
 Db 94 RESSETPPPKPNVT 109
 |||:||||:
 QY 215 RDCAETTPPKPKLS 230

RESULT 11
 ID VPAP_VZVD STANDARD; PRT; 408 AA.
 AC P09274;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE DNA POLYMERASE PROCESSIVITY FACTOR (POLYMERASE ACCESSORY PROTEIN)
 DE (PAP) (DNA-BINDING GENE 16 PROTEIN).
 GN 16.
 OS Varicella-zoster virus (strain Dumas) (VZV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirus.
 RN [1]
 RP SEQUENCE FROM N.A.

```

FT DOMAIN 525 732 MCM.
FT NP_BIND 575 582 ATP (POTENTIAL).
SQ SEQUENCE 1017 AA; 112952 MW; 6DA86379ABDDF774 CRC64;

Query Match 34.2%; Score 63; DB 1; Length 1017;
Best Local Similarity 35.0%; Pred. No. 1.44e+00;
Matches 7; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Db 356 WOKVRIQENANEIPTGSMR 375
QY 209 WQGEIRDACAETPTPKPK 228
||: ||: ||: ||:
||: ||: ||: ||:

RESULT 13
ID POP1_BOVIN STANDARD; PRT; 48 AA.
AC P01154;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 21-JUL-1986 (Rel. 01, Last annotation update)
DE POSTERIOR PITUITARY PEPTIDE.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Butheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE.
RX MEDLINE; 66045180.
RA Preddie E.C.;
RT "Structure of a large polypeptide of bovine posterior pituitary
   tissue.";
RL J. Biol. Chem. 240:4194-4203(1965).
DR FIR; A01418; POBO.
FT DISULFID 34 37
FT DISULFID 40 45
SQ SEQUENCE 48 AA; 5080 MW; 825757EE1567607E CRC64;

Query Match 33.7%; Score 62; DB 1; Length 48;
Best Local Similarity 52.9%; Pred. No. 2.20e+00;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 1 RGEVKDASGELEPPPG 17
QY 211 RGEIRDACAETPTPKPK 227
|||: ||: ||: ||:
|||: ||: ||: ||:

RESULT 14
ID VG04_HSV11 STANDARD; PRT; 179 AA.
AC Q00108;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE HYPOTHETICAL GENE 4 PROTEIN.
GN 4.
OS Itatuturid herpesvirus 1 (Channel catfish virus) (CCV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC unclassified Herpesviridae.
[1]
RN RN
RP RN
RC STRAIN=AUBURN 1;
RX MEDLINE; 92087490.
RA Davison A.J.;
RT "Channel catfish virus: a new type of herpesvirus.";
RL Virology 186:9-14(1992).
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement
CC or send an email to license@isb-sib.ch).
CC CC
CC EMBL; M75136; AAA88185.1;

```

```
DR EMBL; W75136; AAA88107.1; -.
DR FIR; E36786; E36786.
KW Hypothetical protein.
SQ SEQUENCE 179 AA; 19644 MW; 94FAC3315EF229E6 CRC64;

Query Match
Best Local Similarity 33.7%; Score 62; DB 1; Length 179;
Matches 9; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

Db 68 EVDGACSEPTPRPEPRF 85
   | : ||| ||| | :
QY 213 EIRDACAETPT-PKPKL 229

RESULT 15
ID YCAJ_PLAFA STANDARD; PRT; 252 AA.
AC P25407;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE HYPOTHETICAL PROTEIN IN CALMODULIN 5' REGION (FRAGMENT).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91304518.
RA Robson K.J., Jennings M.W.;
RT "The structure of the calmodulin gene of Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 46:19-34(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M59770; AAA29511.1; -.
DR FIR; A45594; A45594.
KW Hypothetical protein.
FT NON_TER 252
SQ SEQUENCE 252 AA; 29324 MW; 5283E694BCCDDF3C CRC64;

Query Match
Best Local Similarity 58.3%; Score 62; DB 1; Length 252;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 74 WPKGYIRKGCAE 85
   | : ||| ||| | :
QY 209 WQGEIRDCAE 220

Search completed: Wed May 10 13:35:18 2000
Job time : 98 secs.
```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 13:35:36 2000; MasPar time 238.82 Seconds
Tabular output not generated. 6.968 Million cell updates/sec

Title: >US-09-376-430-2
Description: (208-231) from US09376430A.pep (19 of 25)
Perfect Score: 184
Sequence: 1 CWRGEIRDACAETPTPPKPKLSK 24

Scoring table:
PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrmb112
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 29.340; Variance 41.837; scale 0.701

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	75	40.8	2274	11	APC2 PROTEIN.	3.67e-02
2	73	39.7	516	4	PIG10.	8.91e-02
3	69	37.5	79	2	ORF5.	5.05e-01
4	68	37.0	77	14	ENVELOPE GLYCOPROTEIN, 7.73e-01	
5	67	36.4	233	2	HYPOTHETICAL 24.6 KD P	1.18e+00
6	67	36.4	610	11	ENVELOPE GLYCOPROTEIN, 7.73e-01	
7	67	36.4	1551	11	ENVELOPE GLYCOPROTEIN, 7.73e-01	
8	67	36.4	1557	5	ENVELOPE GLYCOPROTEIN, 7.73e-01	
9	66	35.9	605	3	ENVELOPE GLYCOPROTEIN, 7.73e-01	
10	66	35.9	733	2	ENVELOPE GLYCOPROTEIN, 7.73e-01	
11	66	35.9	876	10	ENVELOPE GLYCOPROTEIN, 7.73e-01	
12	65	35.3	228	5	ENVELOPE GLYCOPROTEIN, 7.73e-01	
13	65	35.3	469	3	ENVELOPE GLYCOPROTEIN, 7.73e-01	
14	65	35.3	471	5	ENVELOPE GLYCOPROTEIN, 7.73e-01	
15	65	35.3	644	2	ENVELOPE GLYCOPROTEIN, 7.73e-01	
16	65	35.3	878	2	ENVELOPE GLYCOPROTEIN, 7.73e-01	
17	65	35.3	1347	2	ENVELOPE GLYCOPROTEIN, 7.73e-01	
18	64	34.8	315	14	ENVELOPE GLYCOPROTEIN, 7.73e-01	
19	64	34.8	315	14	ENVELOPE GLYCOPROTEIN, 7.73e-01	
20	64	34.8	985	2	ENVELOPE GLYCOPROTEIN, 7.73e-01	

21	63	34.2	80	14	Q98128	CORE PROTEIN AND MEMBR	6.07e+00
22	63	34.2	106	1	Q9YB11	106AA LONG HYPOTHETICA	6.07e+00
23	63	34.2	165	4	Q75798	ORF-3.	6.07e+00
24	63	34.2	335	5	O02127	COSMID E03H12.	6.07e+00
25	63	34.2	409	14	O83372	V-MOS TRANSFORMING PRO	6.07e+00
26	63	34.2	482	14	O9WR14	ENVELOPE GLYCOPROTEIN	6.07e+00
27	63	34.2	879	11	O9Z2H5	NEURONAL PROTEIN 4.1.	6.07e+00
28	63	34.2	879	11	O9WTP0	RAT BRAIN 4.1(S).	6.07e+00
29	63	34.2	934	4	O15046	KIAA0338 (FRAGMENT).	6.07e+00
30	63	34.2	985	2	O9ZNI6	SERINE PROTEASE HOMOLO	6.07e+00
31	62	33.7	78	14	Q73837	ENVELOPE GLYCOPROTEIN,	9.06e+00
32	62	33.7	78	14	Q73836	ENVELOPE GLYCOPROTEIN,	9.06e+00
33	62	33.7	83	2	O05789	MOAD.	9.06e+00
34	62	33.7	101	1	O9YB5	101AA LONG HYPOTHETICA	9.06e+00
35	62	33.7	131	1	O9YF50	101AA LONG HYPOTHETICA	9.06e+00
36	62	33.7	207	5	O02128	HYPOTHETICAL 22.3 KD P	9.06e+00
37	62	33.7	237	5	O02128	COSMID E03H12.	9.06e+00
38	62	33.7	344	10	O22904	HYPOTHETICAL 50.5 KD P	9.06e+00
39	62	33.7	479	2	O9Z5E2	HYPOTHETICAL 50.5 KD P	9.06e+00
40	62	33.7	707	5	O17279	W09B6.2 PROTEIN.	9.06e+00
41	62	33.7	852	2	P76087	HYPOTHETICAL 87.1 KD P	9.06e+00
42	62	33.7	852	2	P76087	RTOA PROTEIN.	9.06e+00
43	62	33.7	1209	4	Q9Y4D3	KIAA0649 PROTEIN.	9.06e+00
44	62	33.7	1277	14	Q08547	TRANS-ACTING TRANSCRIP	9.06e+00
45	61	33.2	1332	5	O45599	H02112.1 PROTEIN.	1.34e+01

ALIGNMENTS

RESULT	1
ID	Q941K7
AC	Q941K7; PRELIMINARY; PRT; 2274 AA.
DT	01-MAY-1999 (TRENBLrel. 10, Created)
DT	01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT	01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE	APC2 PROTEIN.
GN	APC2.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN	[1]
RP	SEQUENCE FROM N.A.
FX	MEDLINE; 99147086.
RA	VAN ES J.H., KIRKPATRICK C., VAN DE WETERING M., MOLENAAR M.,
RA	MILES A., KUIPERS J., DESTREE O., PEIFER M., CLEVERS H.;
RT	"Identification of APC2, a homologue of the adenomatous polyposis coli
RT	tumour suppressor.";
RL	Curr. Biol. 9:105-108(1999).
DR	EMBL; AJ130783; CAAL0207.1; JOINED.
DR	EMBL; AJ130784; CAAL0207.1; JOINED.
DR	EMBL; AJ130785; CAAL0207.1; JOINED.
DR	EMBL; AJ130786; CAAL0207.1; JOINED.
DR	EMBL; AJ130787; CAAL0207.1; JOINED.
DR	EMBL; AJ130788; CAAL0207.1; JOINED.
DR	EMBL; AJ130789; CAAL0207.1; JOINED.
DR	EMBL; AJ130790; CAAL0207.1; JOINED.
DR	EMBL; AJ130791; CAAL0207.1; JOINED.
DR	EMBL; AJ130792; CAAL0207.1; JOINED.
DR	EMBL; AJ130793; CAAL0207.1; JOINED.
DR	EMBL; AJ130794; CAAL0207.1; JOINED.
DR	EMBL; AJ130795; CAAL0207.1; JOINED.
DR	EMBL; AJ130796; CAAL0207.1; JOINED.
SQ	SEQUENCE 2274 AA; 243137 MW; BA29D29A CRC32;

Query Match * 40.8%; Score 75; DB 11; Length 2274;
Best Local Similarity 40.9%; Pred. No. 3.67e-02;
Matches 9; Conservancy 6; Mismatches 7; Indels 0; Gaps 0;

Db 383 CWDWLOARDSGTETPTPTPEI 404

QY 208 CWRGEIRDACAETPTPPKPKL 229

```

RESULT 2
ID O14682 PRELIMINARY; PRT; 516 AA.
AC O14682;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE PIG10.
DE PIG10.
GN PIG10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97449378.
RA POLYAK K., XIA Y., ZWEIER J.L., KINZLER K.W., VOGELSTEIN B.;
RT "A model for p53-induced apoptosis.";
RL Nature 389:300-306(1997).
DR EMBL; AF010314; AAC39532.1; -.
DR PFAM; PF00651; BTB; 1.
DR PFAM; PF01344; Kelch; 3.
SQ SEQUENCE 516 AA; 58314 MW; 62C0B738 CRC32;

Query Match 39.7%; Score 73; DB 4; Length 516;
Best Local Similarity 69.2%; Pred. No. 8.91e-02;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 129 CWSFKDIRDCAE 141
QY 208 CWQGEIRDCAE 220

RESULT 3
ID O34036 PRELIMINARY; PRT; 79 AA.
AC O34036;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE ORF5.
OS Streptococcus thermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CNR21205;
RX MEDLINE: 98048466.
RA STANLEY E., FITZGERALD G.F., LE MARREC C., FAYARD B., VAN SINDEREN D.;
RT "Sequence analysis and characterization of phi O1205, a temperate
bacteriophage infecting Streptococcus thermophilus CNR21205.";
RL Microbiology 143:0-0(0).
DR EMBL; U88974; AAC79521.1; -.
SQ SEQUENCE 79 AA; 9335 MW; 0AA461C5 CRC32;

Query Match 37.5%; Score 69; DB 2; Length 79;
Best Local Similarity 44.4%; Pred. No. 5.05e-01;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

DB 44 CWOQSEISKACELLQIPS 61
QY 208 CWQGEIRDCAETPTTP 225

RESULT 4
ID Q73840 PRELIMINARY; PRT; 77 AA.
AC Q73840;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE ENVELOPE GLYCOPROTEIN, V1-V2 REGION (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviruses; Retroviridae; Lentivirus.
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-PATIENT P1.
RX MEDLINE: 96251940.
RA PALMER C., BALFE P., FOX D., MAY J.C., FREDERIKSSON R., FENYO E.M.,
RA MCKEATING J.A.;
RT "Functional characterization of the VIV2 region of human
immunodeficiency virus type 1.";
RL Virology 220:436-449(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT P1.
RA BLOUIN C.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U52258; AAB07912.1; -.
DR PFAM; PF00516; GPI20; 1.
KW Envelope protein.
FT NON_TER 1
FT NON_TER 77
SQ SEQUENCE 77 AA; 8712 MW; 4227D7E7 CRC32;

Query Match 37.0%; Score 68; DB 14; Length 77;
Best Local Similarity 40.7%; Pred. No. 7.73e-01;
Matches 11; Conservative 7; Mismatches 5; Indels 4; Gaps 3;

DB 18 CWEKMRGEIRN-CSFNVTTSIRNMKK 43
QY 208 CWO--RGEIRDACA-ETPTPKPKLSK 231

RESULT 5
ID O53448 PRELIMINARY; PRT; 233 AA.
AC O53448;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE HYPOTHETICAL 24.6 KD PROTEIN.
GN MTV017.53
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA DEVLIN K., CHURCHER C.M.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA COLE S.T., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE: 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA COLE S.T.;
RT "An integrated map of the genome of the tubercle bacillus.
Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium
leprae.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
KW Hypothetical protein.
SQ SEQUENCE 233 AA; 24563 MW; 2A6CC8D7 CRC32;

Query Match 36.4%; Score 67; DB 2; Length 233;
Best Local Similarity 38.1%; Pred. No. 1.18e-00;
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

DB 15 WDTGHVTAEPQTPRPAPRL 35
QY 209 WQGEIRDACAETPTTPPKL 229

```

```

RESULT 6
ID Q9WVR8 PRELIMINARY; PRT; 610 AA.
AC Q9WVR8
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE MENIN.
GN MENI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=BRAIN;
RA MARYAMA K., TSURADA T., HOSONO T., OHKURA N., KISHI M., HONDA M.,
RA NARA-ASHIZAWA N., NAGASAKI K., YAMAGUCHI K.;
RT "Molecular cloning and tissue distribution of mRNA for rat menin.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB023400; BAA82134.1; -.
SQ SEQUENCE 610 AA; 67335 MW; E957EDE7 CRC32;

Query Match 36.4%; Score 67; DB 11; Length 610;
Best Local Similarity 45.8%; Pred. No. 1.18e+00;
Matches 11; Conservative 5; Mismatches 7; Indels 1; Gaps 1;

Db 479 RRGPRRESKPEPPPPKPAIDK 502
QY 209 WQGEIRDACAETPTPPK-PLSK 231

RESULT 7
ID Q9WTP1 PRELIMINARY; PRT; 1551 AA.
AC Q9WTP1
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE RAT BRAIN 4.1(L).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RA YAMAKAWA H., OHARA R., NAKAJIMA D., NAKAYAMA M., OHARA O.;
RT "Molecular characterization of a new member of the protein 4.1 family
RT (brain 4.1) in rat brain.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB019257; BAA76625.1; -.
DR PROSITE; PS00660; BAND.41.1; 1.
DR PROSITE; PS00661; BAND.41.2; 1.
SQ SEQUENCE 1551 AA; 171010 MW; B742227C CRC32;

Query Match 36.4%; Score 67; DB 11; Length 1551;
Best Local Similarity 36.8%; Pred. No. 1.18e+00;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 1228 QRAGLRGSEKVKPPRP 1246
QY 210 QRGEIRDACAETPTPPK 228

RESULT 8
ID Q96652 PRELIMINARY; PRT; 1557 AA.
AC Q96652
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE CONNECTOR ENHANCER OF KSR PROTEIN CNK.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE; 99029842.
RA THERRIEN M., WONG A. M., RUBIN G. M.;
RT "CNK, a RAF-binding multidomain protein required for RAS signaling.";
RL Cell 95:343-353(1998).
DR EMBL; AF100152; AAC80557.1; -.
SQ SEQUENCE 1557 AA; 171447 MW; 3433F0C3 CRC32;

Query Match 36.4%; Score 67; DB 5; Length 1557;
Best Local Similarity 50.0%; Pred. No. 1.18e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 371 SCSDIPTDPKLA 384
QY 217 ACAETPTPPKLS 230

RESULT 9
ID Q00292 PRELIMINARY; PRT; 605 AA.
AC Q00292
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE DIHYDROGEODIN OXIDASE PRECURSOR (EC 1.10.3.-) (DHGO).
GN DHGO.
OS Aspergillus terreus.
OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes;
OC Eurotiales; Trichocomaceae; anamorphic Trichocomaceae; Aspergillus.
RN [1]
RP SEQUENCE FROM N.A.; SEQUENCE OF 22-31, AND FUNCTION.
RC STRAIN=IMI 16043; TISSUE=MYCELIUM;
RX MEDLINE; 95394900.
RA HUANG K.-X., FUJII I., EBIZUKA Y., GOMI K., SANKAWA U.;
RT "Molecular cloning and heterologous expression of the gene encoding
RT dihydrogeodin oxidase, a multicopper blue enzyme from Aspergillus
RT terreus.";
RL J. Biol. Chem. 270:21495-21502(1995).
RN [2]
RP CHARACTERIZATION OF FUNCTION, COPPER-BINDING, AND SUBUNITS.
RC STRAIN=IMI 16043; TISSUE=MYCELIUM;
RX MEDLINE; 87194683
RA FUJII I., IJIMA H., TSUKITA S., EBIZUKA Y., SANKAWA U.;
RT "Purification and properties of dihydrogeodin oxidase from Aspergillus
RT terreus.";
RL J. Biochem. 101:11-18(1987).
CC -!- FUNCTION: CATALYZES A REGIO- AND STEREOSPECIFIC INTRAMOLECULAR
CC PHENOL OXIDATIVE COUPLING REACTION.
CC -!- CATALYTIC ACTIVITY: 2 DIHYDROGEODIN + O(2) = 2 (+)-GEODIN + H(2)O.
CC -!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU CENTRES
CC KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE 3 OR COUPLED
CC BINUCLEAR (BY SIMILARITY).
CC -!- PATHWAY: THE FORMATION OF (+)-GEODIN FROM ACETYL-COA AND MALONYL-
CC COA.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
DR EMBL; D49538; BAA08486.1; -.
DR MENDEL; 20780; Aspte;3152;20780.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; FALSE_NEG.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
DR PFAM; PF00394; Cu-oxidase; 2.
KW Oxidoreductase; Copper; Metal-binding; Signal; Glycoprotein.
FT SIGNAL 1 21
FT CHAIN 22 605 DIHYDROGEODIN OXIDASE.
FT DOMAIN 59 179 PLASTOCYANIN-LIKE 1.
FT DOMAIN 188 354 PLASTOCYANIN-LIKE 2.
FT DOMAIN 428 561 PLASTOCYANIN-LIKE-3.
FT METAL 117 117 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 119 119 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 161 161 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 163 163 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 484 484 COPPER (TYPE 1) (BY SIMILARITY).

```

```

FT METAL 487 487 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 491 491 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 543 543 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 544 544 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 545 545 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 549 549 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 554 554 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 554 554 COPPER (TYPE 1) (BY SIMILARITY).
FT CARBOHYD 27 27 POTENTIAL.
FT CARBOHYD 107 107 POTENTIAL.
FT CARBOHYD 112 112 POTENTIAL.
FT CARBOHYD 278 278 POTENTIAL.
FT CARBOHYD 467 467 POTENTIAL.
FT CARBOHYD 478 478 POTENTIAL.
SQ SEQUENCE 605 AA; 67549 MW; 020AAC2A CRC32;

Query Match 35.9%; Score 66; DB 3; Length 605;
Best Local Similarity 34.8%; Pred. No. 1.78e+00;
Matches 8; Conservative 5; Mismatches 9; Indels 1; Gaps 1;

Db 37 CWQDGFDSWDYDPKVPAPGKL 59
QY 208 CWQRG-EIRDACAETPTPPKPL 229

RESULT 10
ID Q9XDH2 PRELIMINARY; PRT; 763 AA.
AC Q9XDH2;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PROLINE-RICH MUCIN HOMOLOG.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA ESPITIA C., LACLETTE J.L., MONDRAGON-PALOMINO M., MARTENS A.,
RA ZHANG Y., MORENO C., SINGH M.;
RT "Cloning and characterization of a new member of the PGRS family that
RT is a useful marker of polymorphism in Mycobacterium tuberculosis.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF071081; AAD41594.1; -.
SQ SEQUENCE 763 AA; 75035 MW; A9E577DF CRC32;

Query Match 35.9%; Score 66; DB 2; Length 763;
Best Local Similarity 40.9%; Pred. No. 1.78e+00;
Matches 9; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

Db 38 CWMLVSAAPCPAPPAPPPKPK 59
QY 208 CWQGEIRDACAETPTPPKPK 228

RESULT 11
ID O6483 PRELIMINARY; PRT; 876 AA.
AC O6483;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE HYPOTHETICAL 98.7 KD PROTEIN.
GN T20K24.21.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eucosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROUNSLEY S.D., KAUL S., LIN X., KETCHUM K.A., CROSBY M.L.,
RA BRANDON R.C., SYKES S.M., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;

```

```

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC002392; AAD12037.1; -.
DR MENDEL: 29334; Arath:1197;29334.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PFAM: PF00560; LRR; 2.
DR PFAM: PF00069; pkinase; 1.
KW Hypothetical protein.
SQ SEQUENCE 876 AA; 98711 MW; 1D53F6FD CRC32;

Query Match 35.9%; Score 66; DB 10; Length 876;
Best Local Similarity 36.4%; Pred. No. 1.78e+00;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 815 WKMSIALACTEHTSAQRPTMS 836
QY 209 WQGEIRDACAETPTPPKPLS 230

RESULT 12
ID Q1836 PRELIMINARY; PRT; 228 AA.
AC Q1836; Q19856;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE C54D10.10 PROTEIN.
GN C54D10.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA KERSHAW J.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS J., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA DOBSON R.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z75539; CAA99845.1; -.
DR EMBL: Z75531; CAA99845.1; JOINED.
DR EMBL: Z75531; CAA99808.1; -.
DR EMBL: Z75539; CAA99808.1; JOINED.
DR HSP; P12111; IKUN.
DR PFAM: PF00014; Kunitz_BPTI; 2.
DR PRINTS: PR00759; BASICPTASE.
SQ SEQUENCE 228 AA; 25884 MW; 2A621F93 CRC32;

Query Match 35.3%; Score 65; DB 5; Length 228;
Best Local Similarity 52.9%; Pred. No. 2.70e+00;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 75 RFESRDACSDACVPHKP 91
QY 211 RGEIRDACAETPTPPKP 227

RESULT 13
ID Q06920 PRELIMINARY; PRT; 469 AA.

```


AC Q06920;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)
 DE DNA FOR ORFS AND PPR1 GENE.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA SAVILLE S.P., ATKINSON S., JAMIESON L., POCKLINGTON M.J., ORR E.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA LEVIN D.E., STEVENSON W.D.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X90564; CAA62158.1;
 SQ SEQUENCE 469 AA; 54210 MW; 32C2F8C8 CRC32;
 Query Match 35.3%; Score 65; DB 3; Length 469;
 Best Local Similarity 42.9%; Pred. No. 2.70e+00;
 Matches 9; Conservative 4; Mismatches 7; Indels 1; Gaps 1;
 Db 392 WOHGEARDESNDKNTTSARKK 412
 ||||| : |||||
 QY 209 WQGEIRDACAET-PTPKPK 228
 ||||| : |||||
 RESULT 14
 ID Q9XXF9 PRELIMINARY; PRT; 471 AA.
 AC Q9XXF9;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE Y37A1B.5 PROTEIN.
 GN Y37A1B.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA MCMURRAY A.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL; AL023835; CAA19490.1;
 SQ SEQUENCE 471 AA; 52424 MW; 11408992 CRC32;
 Query Match 35.3%; Score 65; DB 5; Length 471;
 Best Local Similarity 39.1%; Pred. No. 2.70e+00;
 Matches 9; Conservative 6; Mismatches 6; Indels 2; Gaps 2;
 Db 328 CWLHGDIRO-Y-DISDPLKVKLN 348
 ||:|:|: :|:|:|:

QY 208 CWQGEIRDACAETPTPKPKLS 230

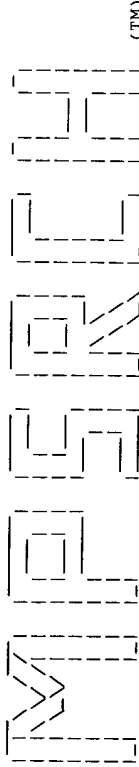
RESULT 15
 ID Q52551 PRELIMINARY; PRT; 644 AA.
 AC Q52551;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE DEHYDROGENASE.
 GN PVAA.
 OS Pseudomonas sp.
 OC Bacteria; Proteobacteria.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MATSUMIYA A., HATAMAKA T., TSUJI M., TAKAKURA K., TAKIZAWA N.,
 RA KIYOHARA H.;
 RT "Molecular Cloning and Nucleotide sequence a Gene that Encoding
 RT Poly(vinyl alcohol) dehydrogenase from Pseudomonas sp. 113P3.";
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D83772; BAA12104.1;
 DR PFAM; PF01011; Bacterial_PQQ; 6.
 DR PFAM; PF00034; Cytochrome_c; 1.
 SQ SEQUENCE 644 AA; 69324 MW; 44C2F3A9 CRC32;

Query Match 35.3%; Score 65; DB 2; Length 644;
 Best Local Similarity 35.3%; Pred. No. 2.70e+00;
 Matches 6; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Db 418 CWQKGREHANCNPPLGP 434

||||| : |||||
 QY 208 CWQGEIRDACAETPTP 224

Search completed: Wed May 10 13:39:46 2000
 Job time : 250 secs.



 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 13:48:52 2000; MasPar time 2.82 Seconds
 Tabular output not generated. 67.122 Million cell updates/sec

Title: >US-09-376-430-2
 Description: (264-271) from US09376430A.pep (20 of 25)
 Perfect Score: 51
 Sequence: 1 SVPDPKSI 8

Scoring table: PAM 150
 Gap 11

Searched: 189963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-geneseq35
 1:geneseqp

Statistics: Mean 14.211; Variance 37.385; scale 0.380

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	41	80.4	359	1 W86316	Kidney injury associat	1.73e+02
2	40	78.4	20	1 W40257	Human wild-type E-Cadh	2.29e+02
3	40	78.4	150	1 R58552	MEAV vaccine.	2.29e+02
4	40	78.4	520	1 W13666	Fragment of dhpA gene	2.29e+02
5	40	78.4	734	1 W13667	Streptomyces viridospo	2.29e+02
6	40	78.4	823	1 W13668	DhpA-mel chimERIC prot	2.29e+02
7	40	78.4	835	1 W85017	GrK5-green fluorescent	2.29e+02
8	40	78.4	842	1 W85008	GrK5-green fluorescent	2.29e+02
9	40	78.4	878	1 R85487	Human E-cadherin precu	2.29e+02
10	40	78.4	878	1 R55060	Sequence of human live	2.29e+02
11	38	74.5	7	1 W15780	Protein kinase C-beta	3.98e+02
12	38	74.5	355	1 R38297	dTDP-D-glucose synthas	3.98e+02
13	38	74.5	671	1 R49533	Type II (beta-1) rat br	3.98e+02
14	38	74.5	672	1 P91373	Protein kinase C beta	3.98e+02
15	38	74.5	673	1 R06957	Sequence of rat brain	3.98e+02
16	38	74.5	673	1 R94764	Type II (beta-2) rat b	3.98e+02
17	38	74.5	680	1 W21708	Cyclomaltodextrin gluc	3.98e+02
18	38	74.5	680	1 P61517	Sequence of cyclomalto	3.98e+02
19	38	74.5	711	1 R42844	Wild type cyclomultode	3.98e+02
20	38	74.5	711	1 R42845	Mutant cyclomultodextr	3.98e+02
21	38	74.5	711	1 R42846	Mutant cyclomultodextr	3.98e+02
22	37	72.5	72	1 Y05318	Human secreted protein	5.22e+02
23	37	72.5	647	1 W73376	Human HPD78 protein	5.22e+02

24	37	72.5	1456	1 R49042	NMDA receptor channel	5.22e+02
25	37	72.5	1463	1 W99482	Murine NCoA-2 protein.	5.22e+02
26	37	72.5	1464	1 W42632	Human transcriptional	5.22e+02
27	37	72.5	1482	1 R44193	Rat NMDA receptor subu	5.22e+02
28	37	72.5	1482	1 R45944	Glutamic acid receptor	5.22e+02
29	37	72.5	1484	1 R80971	Human excitatory amino	5.22e+02
30	37	72.5	1484	1 R56040	Human N-methyl-D-aspar	5.22e+02
31	37	72.5	1484	1 R92507	Human NMDA receptor R2	5.22e+02
32	37	72.5	1484	1 W87510	Human N-methyl-D-aspar	5.22e+02
33	36	70.6	163	1 Y07013	Breast cancer associat	6.84e+02
34	36	70.6	411	1 W95599	S. pneumoniae amps pol	6.84e+02
35	36	70.6	512	1 R37347	Truncated secreted SGF	6.84e+02
36	36	70.6	521	1 W32303	Arabidopsis thaliana i	6.84e+02
37	36	70.6	524	1 W32301	Arabidopsis thaliana i	6.84e+02
38	36	70.6	524	1 W32304	Arabidopsis thaliana i	6.84e+02
39	36	70.6	528	1 W46750	Tomato phosphate trans	6.84e+02
40	36	70.6	771	1 R71380	Human semaphorin III p	6.84e+02
41	36	70.6	776	1 R60178	Lethal factor of Bacil	6.84e+02
42	36	70.6	922	1 W89575	Human DNA ligase III.	6.84e+02
43	36	70.6	1151	1 W77287	Zebrafish differentiat	6.84e+02
44	36	70.6	1978	1 Y07032	Breast cancer associat	6.84e+02
45	36	70.6	3077	1 P93283	Sequence of clone HIV-	6.84e+02

*† ALIGNMENTS

RESULT 1

ID W86316 standard; Protein; 359 AA.
 AC W86316;
 DT 01-MAR-1999 (first entry)
 DE Kidney injury associated molecule HW044 protein.
 KW Kidney injury associated molecule; kidney injury related molecule;
 KW KIM: tissue growth promotion; regeneration; renal condition;
 KW acute renal failure; acute nephritis; tumour.
 OS Rattus sp.
 PN W09853071-A1.
 PD 26-NOV-1998.
 PF 22-MAY-1998; U10547.
 PR 23-MAY-1997; US-047491.
 PR 23-MAY-1997; US-047490.
 PA (BIOJ) BIOGEN INC.
 PI Cate RL, Hession CA, Sanicola-Nadel M, Wei H;
 DR WPI; 99-045312/04.
 DR N-PSDB; V80599.
 PT Kidney injury-associated molecule, KIM, polypeptides - upregulated
 PT in injured or regenerating tissues, useful to promote tissue growth
 PT and regeneration, especially to treat renal conditions
 PS Claim 17; Page 95-96; 213pp; English.
 CC The present sequence represents a kidney injury associated molecule
 CC (KIM) protein. KIM proteins can be administered therapeutically
 CC by expressing KIM encoding polynucleotides, to promote growth and/or
 CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins
 CC are upregulated in injured or regenerating (especially renal) tissues.
 CC KIM fusion proteins, conjugates, antibodies and vectors can also be used
 CC therapeutically, e.g. these or the KIM proteins may be included with an
 CC acceptable carrier in pharmaceutical compositions, useful for therapy/
 CC prophylaxis of conditions associated with dysfunction/disregulation of
 CC KIM genes or proteins, especially renal diseases or impairments of renal
 CC function in humans (e.g. acute renal failure, acute nephritis). The
 CC polynucleotides can be used to produce antisense sequences which, when
 CC internalised into cells, can disrupt expression of a cellular KIM gene,
 CC also useful in therapy (e.g. to block the growth of tumours dependent on
 CC KIM for growth) or compositions. The proteins and polynucleotides are
 CC useful diagnostically e.g. to detect and quantify renal injury/disease
 CC (indicative of increased risk, or presence of, renal injury or impaired
 CC function), or abnormal responses to tissue injury (indicative of
 CC increased risk, or presence of, an autoimmune response or abnormal
 CC tissue growth arising from/affecting renal tissue). The proteins can
 CC also be used to locate KIM-producing cells (especially specific loci,
 CC e.g. tissue masses abnormally producing/expressing KIM such as tumours
 CC arising from/affecting renal tissue), by contacting cells with an
 CC imageable KIM-binding reagent and imaging reagent accumulation.
 SQ Sequence 359 AA;

```

Query Match      80.4%; Score 41; DB 1; Length 359;
Best Local Similarity 71.4%; Pred. No. 1.73e+02;
Matches      5; Conservative      2; Mismatches 0; Indels 0; Gaps 0;

Db      184 SVPEPKA 190
QY      264 SVPDPKS 270

RESULT      2
ID      W40257 standard; Protein; 20 AA.
AC      W40257;
DE      Human wild-type E-cadherin fragment (pos. 588-607).
KW      E-Cadherin; mutant; variant; monoclonal antibody; diagnosis; therapy;
KW      gastric carcinoma; immunoassay; therapeutic conjugate; toxin;
KW      tumour cell.
OS      Homo sapiens.
PN      DE1962938-Cl.
PD      27-NOV-1997.
PF      24-JUL-1996; DE-029938.
PR      (GSF-) GSF-FORSCHUNGSZENTRUM UMWELT & GESUNDHEIT.
PI      Becker K, Eulitz M, Hoefler H, Kremmer E, Schuhmacher C;
PI      WPI; 97-551542/51.
DR      Monoclonal antibodies specific for mutated E-cadherin peptide
PT      sequences - useful for diagnosis and therapy of stomach cancer
PT      Disclosure; Page 7; 23pp; German.
PS      This sequence represents a fragment of the E-cadherin protein which is
CC      used in the construction of the mutant represented in W36611. This
CC      fragment is used in a method where a novel monoclonal antibody recognises
CC      one or more mutated (through deletion or point mutation) E-cadherin
CC      sequences. Such antibodies can be used for diagnosis and therapy of
CC      gastric carcinomas, especially as diagnostic immunoassay reagents or as
CC      therapeutic conjugates with toxins or radionuclides. The oligonucleotides
CC      can be used to detect tumour cells in samples containing human cells. The
CC      oligopeptides can also be used for immunotherapy of tumours, especially
CC      gastric carcinomas.
SQ      Sequence 20 AA;

Query Match      78.4%; Score 40; DB 1; Length 20;
Best Local Similarity 37.5%; Pred. No. 2.29e+02;
Matches      3; Conservative      5; Mismatches 0; Indels 0; Gaps 0;

Db      6 PIPEPRII 13
QY      264 SVPDPKS 271

RESULT      3
ID      R58552 standard; Protein; 150 AA.
AC      R58552;
DE      29-MAR-1995 (first entry)
DE      MEAV vaccine.
KW      HIV-1; V3 loop; multiple epitope; AIDS; vaccine; MEAV;
KW      Escherichia coli; pKK-MEAV.
OS      Synthetic.
FH      Key      Location/Qualifiers
FT      domain      1..29
FT      /label= CD4_binding_domain
FT      peptide      30..35
FT      /label= Spacer
FT      peptide      36..51
FT      /label= MN_peptide
FT      /note= "peptide from V3 loop of HIV-1 isolate MN"
FT      peptide      52..57
FT      /label= Spacer
FT      peptide      58..73
FT      /label= SC_peptide
FT      /note= "peptide from V3 loop of HIV-1 isolate SC"
FT      peptide      74..79
FT      /label= Spacer

```

```

peptide      80..97
FT      /label= RF_peptide
FT      /note= "peptide from V3 loop of HIV-1 isolate RF"
FT      peptide      98..103
FT      /label= Spacer
FT      peptide      104..121
FT      /label= IIIB_peptide
FT      /note= "peptide from V3 loop of HIV-1 isolate IIIB"
FT      peptide      122..127
FT      /label= Spacer
FT      peptide      128..142
FT      /label= WMJ2_peptide
FT      /note= "peptide from V3 loop of HIV-1 isolate WMJ2"
FT      peptide      143..148
FT      /label= Spacer
FT      WO9418234-A.
PN      18-AUG-1994.
PD      10-FEB-1994; U01523.
PR      10-FEB-1993; US-015770.
PA      (UNBI-) UNITED BIOMEDICAL INC.
PI      Shen DF, Wang CY;
PI      WPI; 94-279687/34.
DR      N-PSDB; Q70535.
DR      New recombinant proteins 'ontg multiple antigenic determinants -
PT      linked by flexible hinge domains
PT      Disclosure; Page 37-38; 56pp; English.
PS      MEAV gene encodes a portion of the CD4 binding domain (R58550)
CC      of HIV env protein, the domain being capable of inducing a helper T-
CC      cell response, and 5 peptide domains from the V3 loop of HIV-1
CC      isolates MN, SC, RF, IIIB and WMJ2 (R5845-49), each peptide being
CC      separated by a spacer domain (R58551). The gene was expressed in E.
CC      coli BL21/pKK-MEAV for preparation of a multiple epitope AIDS
CC      vaccine.
SQ      Sequence 150 AA;

Query Match      78.4%; Score 40; DB 1; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.29e+02;
Matches      6; Conservative      0; Mismatches 0; Indels 0; Gaps 0;

Db      77 PDPKSI 82
QY      266 PDPKSI 271

RESULT      4
ID      W13666 standard; protein; 520 AA.
AC      W13666;
DE      06-OCT-1997 (first entry)
DE      Fragment of dhpA gene product.
KW      asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
KW      derivative; Streptomyces viridosporus; ester; chiral; synthesis;
KW      cardiovascular; treatment; hypertension; ischaemic heart disease.
OS      Streptomyces viridosporus.
FH      Key      Location/Qualifiers
FT      misc_difference 29
FT      /note= "Asp29 as indicated in the specification"
FT      misc_difference 238
FT      /note= "Ser238 as indicated in the specification"
FT      WO9705243-A1.
PN      13-FEB-1997.
PD      30-JUL-1996; J02147.
PR      31-JUL-1995; JP-212975.
PR      29-FEB-1996; JP-067478.
PA      (SAOC) MERCIAN CORP.
PI      Arisawa A, Dobashi K, Isshiki K, Matsufuji M, Nakashima T;
PI      Tsuruta T, Yoshiooka T;
PI      WPI; 97-145682/13.
DR      N-PSDB; T61454-55.
DR      Asymmetric hydrolase gene derived from Streptomyces viridosporus -
PT      acts on 4-substituted-1,4-di:hydro:pyridine derivatives to produce
PT      chiral derivatives useful for synthesis of cardiovascular drugs
PS      Claim 1; Page 60-63; 78pp; Japanese.
CC      This sequence is a fragment of an asymmetric hydrolase which acts on

```

CC 4-substituted-1,4-dihydropyridine derivatives. The DNA sequence encoding
 CC the hydrolase (DhpA) was isolated from the chromosomal DNA of
 CC Streptomyces viridosporus. The enzyme allows the efficient conversion
 CC of 4-substituted-1,4-dihydropyridine esters to chiral partially
 CC hydrolysed derivatives, for use in the synthesis of cardiovascular drugs
 CC suitable for the treatment of e.g. hypertension and ischaemic heart
 CC disease.
 SQ Sequence 520 AA;

Query Match 78.4%; Score 40; DB 1; Length 520;
 Best Local Similarity 71.4%; Pred. No. 2.29e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 508 TVPDPKA 514
 QY 264 SVPDPKS 270

RESULT 5
 ID W13667 standard; Protein; 734 AA.

AC W13667;
 DT 06-OCT-1997 (first entry)
 DE Streptomyces viridosporus dhpA gene product.
 KW asymmetric hydrolase; dhpA: 4-substituted-1,4-dihydropyridine;
 KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
 KW cardiovascular; treatment; hypertension; ischaemic heart disease.
 OS Streptomyces viridosporus.
 FH Key Location/Qualifiers
 FT peptide 205..724
 FT /note= "fragment of dhpA protein; see W13666"

PN W09705243-A1.
 PD 13-FEB-1997.
 PF 30-JUL-1996; J02147.
 PR 31-JUL-1995; JP-212975.
 PR 29-FEB-1996; JP-067478.
 PA (SAOC) MERCIAN CORP.
 PI Arisawa A, Dobashi K, Isshiki K, Matsufuji M, Nakashima T;
 PI Tsuruta T, Yoshioka T;
 DR WPI: 97-145682/13.
 DR N-PSDB; T61454.
 PT Asymmetric hydrolase gene derived from Streptomyces viridosporus -
 PT acts on 4-substituted-1,4-di:hydro:pyridine derivatives to produce
 PT chiral derivatives useful for synthesis of cardiovascular drugs
 PS Claim 3; Page 49-55; 78pp; Japanese.
 CC This sequence is an asymmetric hydrolase encoded by the Streptomyces
 CC viridosporus dhpA gene. The enzyme acts on 4-substituted-1,4-
 CC dihydropyridine derivatives. The enzyme allows the efficient conversion
 CC of 4-substituted-1,4-dihydropyridine esters to chiral partially
 CC hydrolysed derivatives, for use in the synthesis of cardiovascular drugs
 CC suitable for the treatment of e.g. hypertension and ischaemic heart
 CC disease.
 SQ Sequence 734 AA;

Query Match 78.4%; Score 40; DB 1; Length 734;
 Best Local Similarity 71.4%; Pred. No. 2.29e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 712 TVPDPKA 718
 QY 264 SVPDPKS 270

RESULT 6

ID W13668 standard; Protein; 823 AA.

AC W13668;
 DT 06-OCT-1997 (first entry)
 DE DhpA-mel chimeric protein.
 KW asymmetric hydrolase; dhpA: 4-substituted-1,4-dihydropyridine;
 KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
 KW cardiovascular; treatment; hypertension; ischaemic heart disease.
 OS Chimeric Streptomyces viridosporus.
 OS Chimeric Streptomyces viridosporus.
 FH Key Location/Qualifiers

FT protein 1..734
 FT /label= DhpA_protein_product
 FT peptide 205..724
 FT /note= "see W13666"
 FT peptide 735..823
 FT /note= "melanin (partial sequence)"

PN W09705243-A1.

PD 13-FEB-1997.

PF 30-JUL-1996; J02147.

PR 31-JUL-1995; JP-212975.

PR 29-FEB-1996; JP-067478.

PA (SAOC) MERCIAN CORP.

PI Arisawa A, Dobashi K, Isshiki K, Matsufuji M, Nakashima T;

PI Tsuruta T, Yoshioka T;

DR WPI: 97-145682/13.

DR N-PSDB; T61455.

PT Asymmetric hydrolase gene derived from Streptomyces viridosporus -
 PT acts on 4-substituted-1,4-di:hydro:pyridine derivatives to produce
 PT chiral derivatives useful for synthesis of cardiovascular drugs
 PS Claim 5; Page 37-43; 78pp; Japanese.

CC This sequence is a fusion protein comprising Streptomyces viridosporus
 CC dhpA gene product (an asymmetric hydrolase) which acts on 4-substituted-
 CC 1,4-dihydropyridine derivatives, and melanin from S. antibioticus. The
 CC DhpA enzyme allows the efficient conversion of 4-substituted-1,4-
 CC dihydropyridine esters to chiral partially hydrolysed derivatives,
 CC for use in the synthesis of cardiovascular drugs suitable for the
 CC treatment of e.g. hypertension and ischaemic heart disease.
 SQ Sequence 823 AA;

Query Match 78.4%; Score 40; DB 1; Length 823;
 Best Local Similarity 71.4%; Pred. No. 2.29e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 712 TVPDPKA 718
 QY 264 SVPDPKS 270

RESULT 7

ID W85017 standard; Protein; 836 AA.

AC W85017;

DT 08-FEB-1999 (first entry)

DE Grk5-green fluorescent protein fusion product.

KW Human; Grk5 gene; fusion protein; green fluorescent protein; GFP;

KW intracellular signalling; chimera.

OS Chimeric - Aequorea victoria.

OS Chimeric - Homo sapiens.

PN W09845704-A2.

PD 15-OCT-1998.

PF 07-APR-1998; DK0145.

PR 07-APR-1997; DK-000392.

PA (NOVO) NOVO-NORDISK AS.

PI Kasper A, Petersen Bjorn S, Scudder K, Thastrup O,

PI Tullin S;

DR WPI: 98-594491/50.

DR N-PSDB; W71032.

PT Determining effect on signalling pathways in live cells from
 PT redistribution of luminophores - specifically fusions of green
 PT fluorescent protein with a signalling component, and new apparatus,
 PT particularly for identifying toxins and potential therapeutic agents
 PS Example 13; Pages 131-133; 326pp; English.

CC The present sequence represents a human Grk5-green fluorescent

CC protein fusion product. The fusion protein is used in an assay

CC that exemplifies the invention. The specification describes how

CC quantitative information about the influence of a molecule on a cellular

CC response is obtained by recording the variation, caused by the molecule,

CC on mechanically intact living cells, in the spatially distributed light

CC emitted from a luminophore present in the cells. The variation in light

CC emission is processed to provide information that correlates spatial

CC distribution to the degree of the molecule. The method is used to

CC identify agents that (in)directly affect intracellular signalling,

CC especially to screen for potential therapeutic agents or toxins, and

CC to identify new drug targets.

```

SQ Sequence 836 AA;
Query Match 78.4%; Score 40; DB 1; Length 836;
Best Local Similarity 57.1%; Pred. No. 2.29e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 466 VPDPRAV 472
QY 265 VDPKSI 271

RESULT 8
ID W85008 standard; Protein; 842 AA.
AC W85008;
DE Grk5-green fluorescent protein fusion product.
KW Human; Grk5 gene; fusion protein; green fluorescent protein; GFP;
KW Intracellular signalling; chimera.
OS Chimeric - Aequorea victoria.
OS Chimeric - Homo sapiens.
PN W09845704-A2.
PD 15-OCT-1998.
PF 07-APR-1998; DK0145.
PR 07-APR-1997; DK-000392.
PA (NOVO) NOVO-NORDISK AS.
PI Kasper A, Petersen Bjorn S, Scudder K, Thastrup O,
PI Tullin S;
DR WPI: 98-594491/50.
DR N-PSDB; V71024.
PT Determining effect on signalling pathways in live cells from
PT redistribution of luminophores - specifically fusions of green
PT fluorescent protein with a signalling component, and new apparatus,
PT particularly for identifying toxins and potential therapeutic agents
PS Example 13; pages 80-81; 326pp; English.
CC The present sequence represents a human Grk5-green fluorescent
CC protein fusion product. The fusion protein is used in an assay
CC that exemplifies the invention. The specification describes how
CC quantitative information about the influence of a molecule on a cellular
CC response is obtained by recording the variation, caused by the molecule,
CC on mechanically intact living cells, in the spatially distributed light
CC emitted from a luminophore present in the cells. The variation in light
CC emission is processed to provide information that correlates spatial
CC distribution to the degree of the molecule. The method is used to
CC identify agents that (indirectly) affect intracellular signalling,
CC especially to screen for potential therapeutic agents or toxins, and
CC to identify new drug targets.
SQ Sequence 842 AA;

Query Match 78.4%; Score 40; DB 1; Length 842;
Best Local Similarity 57.1%; Pred. No. 2.29e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 718 VPDPRAV 724
QY 265 VDPKSI 271

RESULT 9
ID R85487 standard; Protein; 878 AA.
AC R85487;
DE Human E-cadherin precursor.
KW E-cadherin; T-lymphocyte; alpha-E, beta-7 integrin; cell adhesion;
KW autoimmune disease; Crohn disease; psoriasis.
OS Homo sapiens.
PH Key Location/Qualifiers
FT peptide 1..150
FT /label= Sig_peptide 151..702
FT domain /label= Extracellular_domain
FT /note= "The extracellular domain (amino acids 1-552
FT of the mature protein) is the preferred
FT region for generation of peptides of the

```

```

FT domain invention*
FT 703..726 /label= Transmembrane_domain
FT 727..876 /label= Cytoplasmic_domain
PN W09529693-A1.
PD 09-NOV-1995.
PF 03-MAY-1995; U05518.
PR 03-MAY-1994; US-237919.
PA (BGM) BRIGHAM & WOMENS HOSPITAL.
PI Brenner MB, Cepek KL;
DR WPI: 95-392921/50.
DR N-PSDB; T05764.
PT Inhibiting adhesion of T lymphocytes with E-cadherin - useful for
PT isolating agents to treat auto-immune diseases e.g. Crohn's disease,
PT psoriasis, etc.
PS Disclosure; Page 70-75; 103pp; English.
CC The human E-cadherin protein precursor (R85487) is expressed by
CC an cDNA clone (T05764) derived from human liver. The extracellular
CC domain of E-cadherin is used to generate peptides that specifically
CC bind to heterotypic cognates of E-cadherin and which inhibit adhesion
CC of intra-epithelial T-lymphocytes to E-cadherin-expressing epithelial
CC or endothelial cells in vitro or in vivo, thereby modulating
CC mucosal immune responses. Such peptides are also specifically
CC reactive with a monoclonal antibody (E4.6 or E6.1) that binds to
CC E-cadherin and that can inhibit T-cell binding.
SQ Sequence 878 AA;

Query Match 78.4%; Score 40; DB 1; Length 878;
Best Local Similarity 37.5%; Pred. No. 2.29e+02;
Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 589 PIPEPRTI 596
QY 264 SVPDKSI 271

RESULT 10
ID R55060 standard; Protein; 878 AA.
AC R55060;
DE Sequence of human liver E-cadherin.
KW Human epithelial-cadherin; E-cadherin; cell adhesion molecule; CAM;
KW uvomorulin; L-CAM; Cell CAM 120/80.
OS Homo sapiens.
PN W09411401-A.
PD 26-MAY-1994.
PF 16-NOV-1993; U11097.
PR 17-NOV-1992; US-978897.
PA (UYIA) UNIV YALE.
PI Morrow JS, Rimm DL;
DR WPI: 94-183426/22.
DR N-PSDB; Q65487.
PT Purified human E-cadherin protein and nucleic acid - used to
PT develop prods. for diagnosis, prognosis, therapy and prophylaxis
PT of T-cadherin disorders, e.g. malignancies.
PS Claim 1; Page 59-63; 97pp; English.
CC E-cadherin is a cell adhesion molecule that is also known as
CC uvomorulin, L-CAM and Cell CAM 120/80. The DNA encoding hEC was obt'd.
CC by screening normal human liver and hepatocellular carcinoma cDNA
CC libraries and a colonic epithelial cell cDNA library. The following
CC sequences are specifically claimed: AAs 1-878; 151-878; 30
CC sequential AAs from AAs 308-878; AAs 1-150; AAs 178-289; AAs 290-
CC 401; AAs 402-513; AAs 178-513; AAs 151-703; AAs 1-703; AAs 728-878;
CC AAs 704-878; nucleotide sequences comprising nucleotide numbers 116-
CC 2749; 566-2749; 1-1053; 510-2696; 1332-3000; 540-1500; 348-906; 890-
CC 1648; 384-1208; 641-2046; 685-1336; 880-1661; 1199-1742; 1373-1742;
CC 1705-2204; 2458-2775; DNA encoding at least 30 AAs selected from
CC AAs 308-878. The prods. can be used in the diagnosis, prognosis,
CC therapy and prophylaxis of conditions involving improper E-cadherin
CC expression. Suitable dosages for i.v. admin. of a protein are
CC 20-500 mcg/kg body wt.
SQ Sequence 878 AA;

```

Query Match 78.4%; Score 40; DB 1; Length 878;
 Best Local Similarity 37.5%; Pred. No. 2.29e+02;
 Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 589 PIPEPRTI 596
 QY 264 SVPDPKSI 271

RESULT 11
 ID W15780 standard; Peptide; 7 AA.
 AC W15780;
 DT 31-OCT-1997 (first entry)
 DE Protein kinase C-beta peptide beta C2-3 (201-207).
 KW Signal transduction; cell signalling; modulator; immunomodulator;
 KW protein kinase C; receptor for activated kinase C; RACK; PKC-beta;
 KW cognate; graft rejection; autoimmune disease; allergy; asthma;
 KW therapy.
 KW KW
 OS Homo sapiens.
 PN W09714038-A1.
 PD 17-APR-1997. U16195.
 PF 10-OCT-1996; U16195.
 PR 18-JUN-1996; US-665647.
 PR 10-OCT-1995; US-541964.
 PR 31-JAN-1996; US-594447.
 PA (TERR-) TERRAPIN TECHNOLOGIES INC.
 PI Kauvar LM, Mochly-Rosen D, Napolitano EW, Ron D;
 PI Vasquez NJ, Voronova A;
 DR WPI; 97-236030/21.
 PT Identifying a modulator of intracellular signal transduction - by
 PT determining the interaction of a signal generating peptide with the
 PT test substance, allows modulation of the immune system
 PT Example 2; Page 26; 74pp; English.
 CC This sequence is a peptide, designated beta C2-3, that corresponds
 CC to amino acid residues 201-207 in the C2 region of protein kinase
 CC C-beta (PKC-beta). It is unable to interrupt the interaction
 CC of PKC-beta with its cognate receptor for activated kinase C
 CC (RACK). Other PKC isozyme peptides (see W15778-79, W15781,
 CC W15784-85, W17452-78) can be used in a claimed method for
 CC identifying modulators of intracellular signal transduction.
 CC modulating the signal transduction function.
 SQ Sequence 7 AA;

Query Match 74.5%; Score 38; DB 1; Length 7;
 Best Local Similarity 83.3%; Pred. No. 3.98e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 IPDPKS 6
 QY 265 VPDPKS 270

RESULT 12
 ID R38297 standard; Protein; 355 AA.
 AC R38297;
 DT 03-AUG-1993 (first entry)
 DE DTP-D-glucose synthase.
 KW snor; snob; snom; microbial synthesis; actinomycetes; hybrid;
 KW glycosylated; natural products; prods.; sequencing;
 KW secondary metabolite biosynthesis.
 OS Streptomyces nodosus DSM40109.
 PN W09306219-A.
 PD 01-APR-1993.
 PF 15-SEP-1992; E02111.
 PR 18-SEP-1991; DE-130967.
 PA (FARR.) HOECHST AG.
 PI Brau B, Distler J, Grabley S, Piepersberg W, Sichel P,
 PI Stockmann M, Taleghani KM;
 DR WPI; 93-117540/14.
 DR N-PSDB; Q39093.
 PT sec. metabolite biosynthesis genes from Actinomycetes - isolatable
 PT with hybridisation probes using DNA, useful in microbial synthesis

PT of glycosylated and natural prods. in Actinomycetes
 PS Claim 3; Fig 4; 38pp; German.
 CC The sequence is that of the C-terminal sequence of dTDP-D-glucose
 CC synthase which is encoded by the snod gene, a secondary metabolite
 CC biosynthesis gene.
 SQ Sequence 355 AA;

Query Match 74.5%; Score 38; DB 1; Length 355;
 Best Local Similarity 71.4%; Pred. No. 3.98e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 135 VPDPKSI 141
 QY 265 VPDPKSI 271

RESULT 13
 ID R94953 standard; Protein; 671 AA.
 AC R94953;
 DT 07-AUG-1996 (first entry)
 DE Type I (beta-1) rat brain protein kinase C.
 KW Protein kinase C; signal transduction; tumour; diagnosis; therapy.
 OS Rattus Sp.
 PN EP-686695-A1.
 PD 13-DEC-1995.
 PF 26-JUN-1987; 112109.
 PR 27-JUN-1986; JP-149385.
 PR 18-SEP-1986; JP-217944.
 PR 28-NOV-1986; JP-281870.
 PR 25-FEB-1987; JP-040160.
 PA (TAKE) TAKEDA CHEM IND LTD.
 PI Igarashi K, Kurokawa T, Nishizuka Y, Ono Y;
 DR WPI; 96-021913/03.
 DR N-PSDB; Q74001.
 PT New isolated rat protein kinase C - used to develop prods. for the
 PT study, diagnosis, prevention and treatment of diseases involving
 PT abnormal signal transduction
 PS Example 4; Fig 8; 39pp; English.
 CC Type I (beta-1) rat brain protein kinase C (PKC = R94953) performs
 CC the transduction of extracellular signals into cells through the
 CC phosphorylation of proteins. It can be obtcd. on a large scale
 CC by expression of the encoding cDNA sequence (Q74001), derived from
 CC rat brain, in transformed host cells. Type II PKC (R94764),
 CC which is identical to type I up to amino acid 621, and type III PKC
 CC (R94765) were also identified in rat brain. PKC can be used for
 CC prodn. of antibodies, for investigating cellular transduction
 CC mechanisms, for diagnosis of diseases (e.g. tumours) resulting from
 CC an abnormality in signal transduction, and for screening potential
 CC therapeutics.
 SQ Sequence 671 AA;

Query Match 74.5%; Score 38; DB 1; Length 671;
 Best Local Similarity 83.3%; Pred. No. 3.98e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 201 IPDPKS 206
 QY 265 VPDPKS 270

RESULT 14
 ID P91373 standard; protein; 672 AA.
 AC P91373;
 DT 10-MAR-1993 (revised)
 DT 11-JAN-1990. (first entry)
 DE Protein kinase C beta 1.
 KW Protein kinase C beta 1; cellular growth control; tumorigenesis.
 OS Rat.
 PN W08907654-A.
 PD 24-AUG-1989.
 PF 09-FEB-1989; U00462.
 PR 10-FEB-1988; US-154206.
 PA (PROG) Progenics Pharmaceuticals, Inc.

PI Housey GM;
 DR WPI: 89-263721/36.
 DR N-PSDB; N90750.
 PT Screening for protein inhibitors and activators - by examining phenotypic
 PS changes of two cell lines having different protein prodn.
 PS Disclosure: fig 1: 75pp; English.
 CC Protein kinase C (PKC) beta 1 cDNA was sequenced from clone RP58
 CC isolated from a rat brain cDNA library corresp. to clone RP41. PKC is a
 CC Ca2+- and phospholipid dependent ser/thr protein kinase important in
 CC cellular growth control. Molecular wt. = 76.8kD.
 SQ Sequence 672 AA;

Query Match 74.5%; Score 38; DB 1; Length 672;
 Best Local Similarity 83.3%; Pred. No. 3.98e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 201 IPDPKS 206
 QY 265 VPDPS 270

RESULT 15

ID P80697 standard; protein; 673 AA.
 AC P80697;
 DT 24-OCT-1990 (first entry)
 DE Sequence of rat brain protein kinase C type II (beta-2) joined to
 DE N-terminal portion of type I (beta-1)
 KW Rat kinase protein C; enzyme; EC-2.7.1.37; type I (beta-1);
 KW type II (beta-2); cellular signal transduction mechanism.
 OS Rat.
 FH Key.
 FT Location/Qualifiers
 FT Protein
 FT 1..621
 FT /note="Residues 1-621 of type I (beta-1)"
 FT 622..673
 FT /note="Type II (beta-2)"
 FT
 FN EP-251244-A.
 PD 07-JAN-1988.
 PF 26-JUN-1987; 109223.
 PR 27-JUN-1986; JP-149385.
 PR 25-FEB-1987; JP-040160.
 PA (TAKE) Takeda Chemical Ind Kk.
 PI Ono Y, Kurokawa T, Igarashi K, Nishizuka Y;
 DR WPI: 88-001173/01.
 DR N-PSDB; n81279.
 PT New human and rat protein kinase C -
 PT for studying cellular signal transductor mechanisms
 PS Disclosure; Fig 8-1 to 8-5; 32pp; English.
 CC A rat protein kinase C which is a polypeptide comprising
 CC 1-224 of the sequence shown in p80697 which has the same activity; or the
 CC polypeptide which is partially substituted and has the same
 CC activity is claimed. The kinase may be used to investigate cellular
 CC signal transduction mechanisms or in diagnosis of disease conditions
 CC resulting from defects in the cellular signal transduction, eg tumours.
 CC It may also be used as a screening agent in prevention and treatment of
 CC such diseases.
 SQ Sequence 673 AA;

Query Match 74.5%; Score 38; DB 1; Length 673;
 Best Local Similarity 83.3%; Pred. No. 3.98e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 201 IPDPKS 206
 QY 265 VPDPS 270

Search completed: Wed May 10 13:49:00 2000
 Job time : 8 secs.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed May 10 13:49:17 2000; MasPar time 46.20 Seconds
 Tabular output not generated.
 Title: >US-09-376-430-2
 Description: (264-271) from US09376430A.pep (20 of 25)
 Perfect Score: 51
 Sequence: 1 SVPDPKSI 8

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed May 10 13:49:17 2000; MasPar time 46.20 Seconds
 Tabular output not generated.
 Title: >US-09-376-430-2
 Description: (264-271) from US09376430A.pep (20 of 25)
 Perfect Score: 51
 Sequence: 1 SVPDPKSI 8

Scoring table: PAM 150
 Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-issued
 1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 13.336; Variance 36.202; scale 0.368

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	40	78.4	150	1	US-08-015- Sequence 8, Applicatio	1.21e+02
2	40	78.4	590	1	US-08-221- Sequence 14, Applicati	1.21e+02
3	40	78.4	590	3	PCT-US94-1 Sequence 14, Applicati	1.21e+02
4	40	78.4	590	1	US-08-454- Sequence 14, Applicati	1.21e+02
5	40	78.4	878	3	PCT-US95-0 Sequence 2, Applicatio	1.21e+02
6	40	78.4	878	1	US-08-237- Sequence 2, Applicatio	1.21e+02
7	39	76.5	60	2	US-08-968- Sequence 13, Applicati	1.59e+02
8	39	76.5	1674	2	US-08-968- Sequence 12, Applicati	1.59e+02
9	38	74.5	7	2	US-08-665- Sequence 21, Applicati	1.59e+02
10	38	74.5	7	1	US-08-594- Sequence 7, Applicatio	2.10e+02
11	38	74.5	7	1	US-08-541- Sequence 6, Applicatio	2.10e+02
12	38	74.5	292	2	US-08-879- Sequence 2, Applicatio	2.10e+02
13	38	74.5	355	1	US-08-681- Sequence 34, Applicati	2.10e+02
14	38	74.5	355	1	US-08-196- Sequence 34, Applicati	2.10e+02
15	38	74.5	671	4	5266464-2 Patent No. 5266464.	2.10e+02
16	37	72.5	220	1	US-08-033- Sequence 2, Applicatio	2.75e+02
17	37	72.5	220	1	US-08-472- Sequence 2, Applicatio	2.75e+02
18	37	72.5	220	1	US-08-472- Sequence 2, Applicatio	2.75e+02
19	37	72.5	1456	1	US-08-026- Sequence 8, Applicatio	2.75e+02
20	37	72.5	1482	1	US-08-026- Sequence 2, Applicatio	2.75e+02
21	37	72.5	1482	2	US-08-036- Sequence 2, Applicatio	2.75e+02
22	37	72.5	1484	2	US-08-231- Sequence 56, Applicati	2.75e+02
23	36	70.6	289	1	US-08-486- Sequence 56, Applicati	2.75e+02
					Sequence 8, Applicatio	3.61e+02

24	36	70.6	411	2	US-08-933- Sequence 2, Applicatio	3.61e+02
25	36	70.6	477	1	US-08-136- Sequence 2, Applicatio	3.61e+02
26	36	70.6	615	2	US-08-752- Sequence 9, Applicatio	3.61e+02
27	36	70.6	738	1	US-07-985- Sequence 3, Applicatio	3.61e+02
28	36	70.6	771	3	PCT-US94-1 Sequence 54, Applicati	3.61e+02
29	36	70.6	771	1	US-08-121- Sequence 54, Applicati	3.61e+02
30	36	70.6	771	1	US-08-835- Sequence 54, Applicati	3.61e+02
31	36	70.6	771	2	US-09-060- Sequence 54, Applicati	3.61e+02
32	36	70.6	776	3	PCT-US94-0 Sequence 2, Applicatio	3.61e+02
33	36	70.6	776	1	US-08-021- Sequence 2, Applicatio	3.61e+02
34	36	70.6	776	1	US-08-082- Sequence 2, Applicatio	3.61e+02
35	36	70.6	822	2	US-08-474- Sequence 9, Applicatio	3.61e+02
36	36	70.6	822	2	US-08-474- Sequence 9, Applicatio	3.61e+02
37	36	70.6	822	2	US-08-472- Sequence 8, Applicatio	3.61e+02
38	36	70.6	884	4	5208144-8 Patent No. 5208144.	3.61e+02
39	36	70.6	922	2	US-08-464- Sequence 2, Applicatio	3.61e+02
40	36	70.6	3077	4	5223423-2 Patent No. 5223423.	3.61e+02
41	35	68.6	285	2	US-08-418- Sequence 17, Applicati	4.71e+02
42	35	68.6	398	2	US-08-939- Sequence 6, Applicatio	4.71e+02
43	35	68.6	474	1	US-08-459- Sequence 4, Applicatio	4.71e+02
44	35	68.6	498	1	US-08-373- Sequence 2, Applicatio	4.71e+02
45	35	68.6	498	2	US-08-348- Sequence 2, Applicatio	4.71e+02

ALIGNMENTS

RESULT 1
 ID US-08-015-770B-8 STANDARD; PRT: 150 AA.
 XX
 AC xxxxxx
 AD
 DT
 XX

Sequence 8, Application US/08015770B

Sequence 8, Application US/08015770B
 Patent No. 5683695
 GENERAL INFORMATION:

APPLICANT: Shen, De Fen
 APPLICANT: Wang, Chang Yi
 TITLE OF INVENTION: Production of recombinant proteins
 TITLE OF INVENTION: containing multiple antigenic determinants linked by
 TITLE OF INVENTION: flexible domains
 NUMBER OF SEQUENCES: 73
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: United Biomedical, Inc.
 STREET: 25 Davids Drive
 CITY: Hauppauge
 STATE: NY
 COUNTRY: USA
 ZIP: 11788

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/015,770B
 FILING DATE: 10-FEB-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Wilson, M. Lisa
 REGISTRATION NUMBER: 34,045
 REFERENCE/DOCKET NUMBER: 2002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (516)273-2828
 TELEFAX: (516)273-1717
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 150 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide

DE Sequence 14, Application US/08454439
XX
CC Sequence 14, Application US/08454439
CC Patent No. 5591618
CC GENERAL INFORMATION:
CC APPLICANT: Chantry, David
CC APPLICANT: Gray, Patrick W.
CC APPLICANT: Hookstra, Merle F.
CC TITLE OF INVENTION: A No. 5591618el G Protein-Coupled Receptor
CC TITLE OF INVENTION: Kinase GRK6
CC NUMBER OF SEQUENCES: 24
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
CC ADDRESSEE: Borun
CC STREET: 6300 Sears Tower, 233 South Wacker Drive
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60606
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC FILING DATE: 30-MAY-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/454,439
CC FILING DATE: 08/221,817
CC FILING DATE: 31-MAR-1994
CC APPLICATION NUMBER: 08/123,932
CC FILING DATE: 17 SEP 1993
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: No. 5591618and, Greta E.
CC REGISTRATION NUMBER: 35,302
CC REFERENCE/DOCKET NUMBER: 31981
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (312) 474-6300
CC TELEFAX: (312) 474-0448
CC TELEX: 25-3856
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 590 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 590 AA; 67786 MW; 1788463 CN;

Query Match 78.4%; Score 40; DB 1; Length 590;
Best Local Similarity: 57.1%; Pred. No. 1.21e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

DB 466 VDPRAV 472
QY 265 VDPKSI 271
|||||:
RESULT 5
ID PCT-US95-05518-2 STANDARD; PRT: 878 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 2, Application PC/TUS9505518
CC
CC Sequence 2, Application PC/TUS9505518
CC GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
CC E-cadherin Interactions with T Lymphocytes

CC NUMBER OF SEQUENCES: 8
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Wolf, Greenfield & Sacks P.C.
CC STREET: 600 Atlantic Avenue
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/05518
CC FILING DATE: herewith
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/237,919
CC FILING DATE: 3 May 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Plummer, Elizabeth R.
CC REGISTRATION NUMBER: 36,637
CC REFERENCE/DOCKET NUMBER: B0801/7023
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-720-3500
CC TELEFAX: 617-720-2441
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 878 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 878 AA; 96800 MW; 3948339 CN;

Query Match 78.4%; Score 40; DB 3; Length 878;
Best Local Similarity 37.5%; Pred. No. 1.21e+02;
Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

DB 589 PIPEPRTI 596
QY 264 SVDPKSI 271
::|::|:
RESULT 6
ID US-08-237-919-2 STANDARD; PRT: 878 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 2, Application US/08237919
DE Patent No. 5610281
XX GENERAL INFORMATION:
XX APPLICANT: Brenner, Michael B
XX APPLICANT: Cepek, Karyn L
XX TITLE OF INVENTION: Methods and Compositions for
XX Modulating Heterotypic E-cadherin Interactions with T L
XX NUMBER OF SEQUENCES: 2
XX CORRESPONDENCE ADDRESS:
XX ADDRESSEE: Wolf, Greenfield & Sacks P.C.
XX STREET: 600 Atlantic Avenue
XX CITY: Boston
XX STATE: MA
XX COUNTRY: USA
XX ZIP: 02210
XX COMPUTER READABLE FORM:
XX MEDIUM TYPE: Floppy disk
XX COMPUTER: IBM PC compatible
XX OPERATING SYSTEM: PC-DOS/MS-DOS
XX SOFTWARE: PatentIn Release #1.0, Version #1.25
XX CURRENT APPLICATION DATA:

```
CC APPLICATION NUMBER: US/08/237,919
CC FILING DATE:
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Plumer, Elizabeth R
CC REGISTRATION NUMBER: 36,637
CC REFERENCE/DOCKET NUMBER: B0801/7023
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-720-3500
CC TELEFAX: 617-720-2441
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 878 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 878 AA; 96800 MW; 3948339 CN;

Query Match 78.4%; Score 40; DB 1; Length 878;
Best Local Similarity 37.5%; Pred. No. 1.21e+02;
Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 589 PIPEPRTI 596
QY 264 SVPDPKSI 271

RESULT 7
ID US-08-968-542C-13 STANDARD; PRT; 60 AA.
XX
AC xxxxxx
XX
DT
XX
XX
DE Sequence 13, Application US/08968542C
XX
XX Sequence 13, Application US/08968542C
CC Patent No. 5981728
CC GENERAL INFORMATION:
CC APPLICANT: Myers, et al.
CC TITLE OF INVENTION: dull1 Codes For A No. 5981728el Starch
CC NUMBER OF SEQUENCES: 35
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: McGregor & Adler, LLP
CC STREET: 8011 Candle Lane
CC CITY: Houston
CC STATE: TX
CC COUNTRY: USA
CC ZIP: 77071
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5 floppy disk
CC COMPUTER: Apple Macintosh
CC OPERATING SYSTEM: Macintosh
CC SOFTWARE: Microsoft Word 6.0.1 for Macintosh
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/968,542C
CC FILING DATE: No. 5981728ember 12, 1997
CC CLASSIFICATION: 800
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Benjamin Aaron Adler, Ph.D., J.D.
CC REGISTRATION NUMBER: 35,423
CC REFERENCE/DOCKET NUMBER: D6036
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (713) 777-6908
CC TELEFAX: (713) 777-6908
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 60 amino acid residues
CC TYPE: amino acid

CC APPLICATION NUMBER: US/08/237,919
CC FILING DATE:
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Plumer, Elizabeth R
CC REGISTRATION NUMBER: 36,637
CC REFERENCE/DOCKET NUMBER: B0801/7023
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-720-3500
CC TELEFAX: 617-720-2441
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 878 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 878 AA; 96800 MW; 3948339 CN;

Query Match 76.5%; Score 39; DB 2; Length 60;
Best Local Similarity 57.1%; Pred. No. 1.59e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 45 IPEPKKI 51
QY 265 VPDPKSI 271

RESULT 8
ID US-08-968-542C-12 STANDARD; PRT; 1674 AA.
XX
AC xxxxxx
XX
DT
XX
XX
DE Sequence 12, Application US/08968542C
XX
XX Sequence 12, Application US/08968542C
CC Patent No. 5981728
CC GENERAL INFORMATION:
CC APPLICANT: Myers, et al.
CC TITLE OF INVENTION: dull1 Codes For A No. 5981728el Starch
CC NUMBER OF SEQUENCES: 35
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: McGregor & Adler, LLP
CC STREET: 8011 Candle Lane
CC CITY: Houston
CC STATE: TX
CC COUNTRY: USA
CC ZIP: 77071
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5 floppy disk
CC COMPUTER: Apple Macintosh
CC OPERATING SYSTEM: Macintosh
CC SOFTWARE: Microsoft Word 6.0.1 for Macintosh
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/968,542C
CC FILING DATE: No. 5981728ember 12, 1997
CC CLASSIFICATION: 800
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Benjamin Aaron Adler, Ph.D., J.D.
CC REGISTRATION NUMBER: 35,423
CC REFERENCE/DOCKET NUMBER: D6036
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (713) 777-6908
CC TELEFAX: (713) 777-6908
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1674 amino acid residues
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC DESCRIPTION: amino acid
CC HYPOTHETICAL: no
CC ANTI-SENSE: no
CC FRAGMENT TYPE:
CC SEQUENCE 1674 AA; 188310 MW; 14230240 CN;

Query Match 76.5%; Score 39; DB 2; Length 1674;
Best Local Similarity 57.1%; Pred. No. 1.59e+02;
```

Sequence 6, Application US/08541964

Sequence 6, Application US/08541964

```

CC ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.
CC STREET: 1100 New York Ave, N.W., Suite 600
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: U.S.A.
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/879,260
CC FILING DATE: 19JUN1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/020,204
CC FILING DATE: 20JUN1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Ludwig, Stevan R.
CC REGISTRATION NUMBER: 36,203
CC REFERENCE/DOCKET NUMBER: 0809.4260001/JAG/SRL
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-371-2600
CC TELEFAX: 202-371-2440
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 292 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 292 AA; 33526 MW; 386625 CN;
CC
CC Query Match 74.5%; Score 38; DB 2; Length 292;
CC Best Local Similarity 71.4%; Pred.No. 2.10e+02;
CC Matches 5; Conservative 2; Mismatches 0; Indels
CC
CC Db 32 PVPDPKN 38
CC QY 264 SVPDPKS 270
CC
CC RESULT 13
CC ID US-08-681-953-34 STANDARD; PRT; 355 AA.
CC XX
CC XX
CC AC
CC XX
CC DT
CC XX
CC DE
CC DE
CC XX
CC XX
CC Sequence 34, Application US/08681953
CC
CC Sequence 34, Application US/08681953
CC Patent No. 5710032
CC GENERAL INFORMATION:
CC APPLICANT: Piepersberg, Wolfgang
CC APPLICANT: Stockmann, Michael
CC APPLICANT: Taleghani, Kampliz Mansouri
CC APPLICANT: Distler, Jurgen
CC APPLICANT: Grabley, Susanne
CC APPLICANT: Sichel, Petra
CC APPLICANT: Brau, Barbara
CC TITLE OF INVENTION: Secondary-Metabolite Biosynthesis
CC TITLE OF INVENTION: From Actinomycetes, Method of Iso
CC TITLE OF INVENTION: Use.
CC NUMBER OF SEQUENCES: 34
CC CORRESPONDENCE ADDRESS:
CC ADDRESS: Finnegan, Henderson, Farabow, Garrett &
CC ADDRESSEE: Bunner
CC STREET: 1300 I Street, N.W.
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: United States
CC ZIP: 20005-3315

```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Wed May 10 13:48:27 2000; MasPar time 3.99 Seconds
Tabular output not generated. 94.475 Million cell updates/sec

Title: >US-09-376-430-2
Description: (264-271) from US9376430A.pap (20 of 25)
Sequence: 1 SVPDPKSI 8

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p162
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 20.124; Variance 21.394; scale 0.941

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	46	90.2	558	2	S57953	1.59e+00
2	46	90.2	779	2	S64011	1.59e+00
3	45	88.2	323	2	T05478	2.79e+00
4	43	84.3	112	2	A46717	8.34e+00
5	43	84.3	220	2	T02014	8.34e+00
6	43	84.3	331	2	T06227	8.34e+00
7	43	84.3	503	2	T04673	8.34e+00
8	42	82.4	148	2	E69032	1.42e+01
9	42	82.4	642	2	E72215	1.42e+01
10	42	82.4	1055	2	T05663	1.42e+01
11	42	82.4	1556	2	D36793	1.42e+01
12	41	80.4	95	2	S35510	2.41e+01
13	41	80.4	95	1	XLPF2	2.41e+01
14	41	80.4	332	2	T10790	2.41e+01
15	41	80.4	994	2	A48849	2.41e+01
16	41	80.4	994	2	A32792	2.41e+01
17	41	80.4	997	2	B31981	2.41e+01
18	41	80.4	997	2	S34444	2.41e+01
19	41	80.4	997	2	B31982	2.41e+01
20	41	80.4	997	2	S04651	2.41e+01
21	41	80.4	997	2	A40812	2.41e+01
22	41	80.4	997	1	PWRBSC	2.41e+01
23	41	80.4	1001	1	PWRBFC	2.41e+01

24	41	80.4	1003	2	S07526	Ca2+-transporting ATP	2.41e+01
25	41	80.4	1042	1	A33881	Ca2+-transporting ATP	2.41e+01
26	41	80.4	1042	1	PWRBMC	Ca2+-transporting ATP	2.41e+01
27	41	80.4	1042	2	S04652	Ca2+-transporting ATP	2.41e+01
28	41	80.4	1042	2	A31981	Ca2+-transporting ATP	2.41e+01
29	41	80.4	1042	2	B40812	Ca2+-transporting ATP	2.41e+01
30	41	80.4	1043	2	A31982	Ca2+-transporting ATP	2.41e+01
31	41	80.4	2844	2	S28291	hypothetical protein	2.41e+01
32	40	78.4	151	2	B64470	formate hydrogenlyase	4.04e+01
33	40	78.4	229	2	A69762	hypothetical protein	4.04e+01
34	40	78.4	298	2	S50735	hypothetical protein	4.04e+01
35	40	78.4	590	2	A48277	G protein-coupled rec	4.04e+01
36	40	78.4	590	1	S43372	G protein-coupled rec	4.04e+01
37	40	78.4	728	2	S48569	hypothetical protein	4.04e+01
38	40	78.4	882	1	IUHUCE	cadherin 1 precursor	4.04e+01
39	40	78.4	895	2	T02597	hypothetical protein	4.04e+01
40	40	78.4	999	2	S72267	Ca2+-transporting ATP	4.04e+01
41	40	78.4	999	2	A34307	Ca2+-transporting ATP	4.04e+01
42	40	78.4	1002	2	A36691	Ca2+-transporting ATP	4.04e+01
43	40	78.4	1103	2	T04617	hypothetical protein	4.04e+01
44	40	78.4	1163	2	J60366	tight junction protei	4.04e+01
45	39	76.5	1736	2	A47747	tight junction protei	6.71e+01

ALIGNMENTS

RESULT 1
ENTRY S57953 #type complete
TITLE C4BP protein alpha chain precursor - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_Change 20-Aug-1999

ACCESSIONS S57953
REFERENCE S57953
#authors Hillarp, A.; Thern, A.; Dahlback, B.
#submission Submitted to the EMBL Data Library, July 1995
#description Molecular cloning of rat C4b-binding protein alpha- and beta-chains: structural and functional relationships among human, bovine, rabbit, mouse and rat proteins.

#accession S57953
#status Preliminary
#molecule_type mRNA
#residues 1-558 #label HIL
#cross-references EMBL:250051; NID:g899379; PIDN:CAA90391.1;

CLASSIFICATION #superfamily C4b-binding protein alpha chain; complement factor H repeat homology
PID:g899380

FEATURE
15-72
77-134
139-199
204-258
263-324
328-386
390-443
447-501
#domain complement factor H repeat homology #label FH1
#domain complement factor H repeat homology #label FH2
#domain complement factor H repeat homology #label FH3
#domain complement factor H repeat homology #label FH4
#domain complement factor H repeat homology #label FH5
#domain complement factor H repeat homology #label FH6
#domain complement factor H repeat homology #label FH7
#domain complement factor H repeat homology #label FH8
#length 558 #molecular-weight 62265 #checksum 9928

SUMMARY
Query Match 90.2%; Score 46; DB 2; Length 558;
Best Local Similarity 75.0%; Pred. No. 1.59e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 330 PVPDPKSI 337
QY 264 SVPDPKSI 271
#|||||

RESULT 2
ENTRY S64011 #type complete
TITLE 3-isopropylmalate dehydratase (EC 4.2.1.33) - yeast
ORGANISM #formal_name Saccharomyces cerevisiae

```

DATE      17-May-1996 #sequence_revision 17-May-1996 #text_change
ACCESSIONS S64011; S15039; A24105; S31555
REFERENCE   S64003
#authors    Hebling, U.; Hofmann, B.; Delius, H.
#submission submitted to the Protein Sequence Database, May 1996
#accession  S64011
#molecule_type DNA
#residues   1-779 #label HEB
##cross-references EMBL:72531; NID:g1322462; PID:e243316; PID:g1322463;
MIPS:YGL009c
##experimental_source strain S288C
REFERENCE   S15039
#authors    Skala, J.; Capieaux, E.; Balzi, E.; Chen, W.; Goffeau, A.
#journal    Yeast (1991) 7:281-285
#title      Complete sequence of the Saccharomyces cerevisiae LEU1 gene
            encoding isopropylmalate isomerase.
#cross-references MUID:91353082
#accession  S15039
#status     not compared with conceptual translation
#molecule_type DNA
#residues   1-290, 'TLKH', 292-422, 'M', 424-458, 'I', 460-743, 'K', 745-779
##cross-references GB:S57886; NID:g234317; PID:g234318
##experimental_source strain IL425-2B
REFERENCE   A24105
#authors    Hsu, Y.P.; Schimmel, P.
#journal    J. Biol. Chem. (1984) 259:3714-3719
#title      Yeast LEU1. Repression of mRNA levels by leucine and
            relationship of 5'-noncoding region to that of LEU2.
#cross-references MUID:84162042
#accession  A24105
#molecule_type DNA
#residues   1-48 #label HSU
##cross-references EMBL:K01969; NID:g171834; PID:g171835
REFERENCE   S15040
#authors    Chen, W.; Balzi, E.; Capieaux, E.; Choder, M.; Goffeau, A.
#journal    Yeast (1991) 7:287-299
#title      The DNA sequencing of the 17 kb HindIII fragment spanning the
            LEU1 and ATE1 loci on chromosome VII from Saccharomyces
            cerevisiae reveals the PDR6 gene, a new member of the
            genetic network controlling pleiotropic drug resistance.
#cross-references MUID:91353083
#accession  S31555
#status     translation not shown
#molecule_type DNA
#residues   740-743, 'K', 745-779 #label CHE
##cross-references GB:S58126
GENETICS   SGD:LEU1
#gene       SGD:LEU1
#map_position 7L
#complex     monomer
#keywords    carbon-oxygen lyase; hydro-lyase; monomer; transmembrane
            protein
FEATURE     197-216 #domain transmembrane #status predicted #label TM1N
            466-476 #domain transmembrane #status predicted #label TM2
            #length 779 #molecular-weight 85794 #checksum 9891
SUMMARY
Query Match 90.2%; Score 46; DB 2; Length 779;
Best Local Similarity 75.0%; Pred. No. 1.59e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 316 SVPDPKNV 323
Qy 264 SVPDPKSI 271
RESULT 3
ENTRY T05478# #type complete
TITLE peroxidase (EC 1.11.1.7) prx1 - Arabidopsis thaliana
ALTERNATE_NAMES protein T805.170

```

```

ORGANISM   #formal_name Arabidopsis thaliana #common_name mouse-ear
            cress
DATE      23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
ACCESSIONS T05478
REFERENCE   Z15417
#authors    Bevan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Mewes,
            H.W.; Mayer, K.F.X.; Schueller, C.
#submission submitted to the Protein Sequence Database, February 1998
#accession  T05478
#molecule_type DNA
#residues   1-323 #label BEV
##cross-references EMBL:AL021890
##experimental_source cultivar Columbia; BAC clone T805
GENETICS
#gene       prx1
#map_position 4
#introns    69/3; 132/3; 188/1
#note       T805.170
CLASSIFICATION #superfamily peroxidase
KEYWORDS      oxidoreductase
SUMMARY       #length 323 #molecular-weight 36634 #checksum 4506
Query Match 88.2%; Score 45; DB 2; Length 323;
Best Local Similarity 62.5%; Pred. No. 2.79e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 227 SVPDPKAV 234
Qy 264 SVPDPKSI 271
RESULT 4
ENTRY A46717 #type complete
TITLE colipase precursor, pancreatic - dog
ORGANISM #formal_name Canis lupus familiaris #common_name dog
DATE     21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change
            20-Aug-1999
ACCESSIONS A46717; S11436
REFERENCE   A46717
#authors    Fukuoaka, S.; Zhang, D.E.; Taniguchi, Y.; Scheele, G.A.
#journal    J. Biol. Chem. (1993) 268:11312-11320
#title      Structure of the canine pancreatic colipase gene includes two
            protein-binding sites in the promoter region.
#cross-references MUID:93266588
#accession  A46717
#status     preliminary
#molecule_type nucleic acid
#residues   1-112 #label FUK
##cross-references GB:M63427; NID:g163925; PIDN:AAA03513.1; PID:g163926
            sequence extracted from NCBI backbone (NCBIN:132835,
            NCBIP:132836)
#note
REFERENCE   S11436
#authors    Fukuoaka, S.I.; Taniguchi, Y.; Kitagawa, Y.; Scheele, G.
#journal    Nucleic Acids Res. (1990) 18:5549
#title      Full length cDNA sequence encoding canine pancreatic
            colipase.
#cross-references MUID:91016846
#accession  S11436
#status     preliminary
#molecule_type mRNA
#residues   1-112 #label FU2
##cross-references EMBL:X53564; NID:g841; PIDN:CAA37636.1; PID:g842
CLASSIFICATION #superfamily colipase
SUMMARY       #length 112 #molecular-weight 12035 #checksum 9759
Query Match 84.3%; Score 43; DB 2; Length 112;
Best Local Similarity 62.5%; Pred. No. 8.34e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 17 AVDPDPKGI 24
Qy 264 SVPDPKSI 271

```

```

RESULT      5
ENTRY       T02014      #type complete
TITLE       hypothetical protein T15B16.1 - Arabidopsis thaliana
ORGANISM    #formal_name Arabidopsis thaliana #common_name mouse-ear
            cress
DATE        26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change
            24-Mar-1999
ACCESSIONS  T02014
REFERENCE   Z14488
#authors    Stoneking, T.; Smith, R.
#submission submitted to the EMBL Data Library, November 1998
#description The sequence of A. thaliana T15B16.
#accession  T02014
#status     translated from GB/EMBL/DBJ
#molecule_type DNA
#residues   1-220 #label STO
#cross-references EMBL:AF104919; NID:g3859590; PID:g3859595
#experimental_source cultivar Columbia
GENETICS
#map_position 4
#introns     70/1: 113/1; 143/3; 166/3; 197/3; 215/1
#note        T15B16.1
SUMMARY     #length 220 #molecular-weight 25251 #checksum 473

Query Match      84.3%; Score 43; DB 2; Length 220;
Best Local Similarity 75.0%; Pred. No. 8.34e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 30 TLPDPKSI 37
QY 264 SVPDPKSI 271

RESULT      6
ENTRY       T06227      #type complete
TITLE       peroxidase (EC 1.11.1.7) 2, cationic - soybean
ORGANISM    #formal_name Glycine max #common_name soybean
DATE        30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change
            26-Aug-1999
ACCESSIONS  T06227
REFERENCE   Z15549
#authors    Boutilier, K.; Gijzen, M.; Kufu, K.; Miki, B.
#submission submitted to the EMBL Data Library, December 1998
#description Cationic peroxidase from immature soybean seeds.
#accession  T06227
#status     translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues   1-331 #label BOU
#cross-references EMBL:AF039027; NID:g3982595; PID:g3982596
#experimental_source cultivar Harosoy 63; seed coat
GENETICS
#gene       Prx2
#classification superfamily peroxidase
#keywords    oxidoreductase
SUMMARY     #length 331 #molecular-weight 37204 #checksum 9411

Query Match      84.3%; Score 43; DB 2; Length 331;
Best Local Similarity 50.0%; Pred. No. 8.34e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 232 ATPDPKAV 239
QY 264 SVPDPKSI 271

RESULT      7
ENTRY       T04673      #type complete
TITLE       hypothetical protein F8D20.190 - Arabidopsis thaliana
ORGANISM    #formal_name Arabidopsis thaliana #common_name mouse-ear
            cress
DATE        23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change

```

```

21-May-1999
ACCESSIONS  T04673
REFERENCE   Z15381
#authors    Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Jesse, T.;
            Heijnen, L.; Vos, P.; Mewes, H.W.; Mayer, K.F.X.;
            Schueller, C.
#submission submitted to the Protein Sequence Database, July 1998
#accession  T04673
#molecule_type DNA
#residues   1-503 #label BEV
#cross-references EMBL:AL031135
#experimental_source cultivar Columbia; BAC clone F8D20
GENETICS
#map_position 4
#introns     309/1; 405/2
#note        F8D20.190
SUMMARY     #length 503 #molecular-weight 56991 #checksum 5695

Query Match      84.3%; Score 43; DB 2; Length 503;
Best Local Similarity 75.0%; Pred. No. 8.34e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 355 TLPDPKSI 362
QY 264 SVPDPKSI 271

RESULT      8
ENTRY       E69032      #type complete
TITLE       formate hydrogenlyase, subunit 7 - Methanobacterium
ORGANISM    thermoautotrophicum (strain Delta H)
DATE        05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
            05-Jun-1998
ACCESSIONS  E69032
REFERENCE   A69000
#authors    Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
            Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
            Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
            Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Viclaire, R.;
            Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwanli, N.; Caruso,
            A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
            McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.;
            Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
            J.; Reeve, J.N.
#journal     J. Bacteriol. (1997) 179:7135-7155
#title       Complete genome sequence of Methanobacterium
            thermoautotrophicum Delta H: functional analysis and
            comparative genomics.
#cross-references MUID:98037514
#accession  E69032
#status     preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues   1-148 #label MTH
#cross-references GB:AE000891; GB:AE000666; NID:g2622345; PID:g2622351
#experimental_source strain Delta H
GENETICS
#gene       MTH1239
SUMMARY     #length 148 #molecular-weight 15691 #checksum 1605

Query Match      82.4%; Score 42; DB 2; Length 148;
Best Local Similarity 37.5%; Pred. No. 1.42e+01;
Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 77 AIPEPKAV 84
QY 264 SVPDPKSI 271

RESULT      9
ENTRY       E72215      #type complete
TITLE       oligopeptide ABC transporter, periplasmic

```

```

oligopeptide-binding protein - Thermotoga maritima (strain
MSB8)
#formal_name Thermotoga maritima
#sequence_revision 11-Jun-1999 #text_change
11-Jun-1999
E72215
#accessions
#authors
Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson,
R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson,
W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek,
J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton,
M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.;
Saizberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
#journal
Nature 399:323-329
#title
Evidence for lateral gene transfer between Archaea and
Bacteria from genome sequence of Thermotoga maritima.
#cross-references MUID:99287316
#accession E72215
#status preliminary
#molecule_type DNA
#residues 1-642 #label ARN
#cross-references GB:AE001813; GB:AE000512; NID:g4982321; PID:g4982324;
TIGR:TM1746
##experimental_source strain MSB8
GENETICS
#gene TM1746
#length 642 #molecular-weight 73659 #checksum 4136
SUMMARY
Query Match 82.4%; Score 42; DB 2; Length 642;
Best Local Similarity 85.7%; Pred. No. 1.42e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 57 VPDPSL 53
QY 265 VPDPSI 271
|||||:

```

```

RESULT 10
ENTRY T05663 #type complete
TITLE hypothetical protein F22I13.120 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
23-Jul-1999
ACCESSIONS T05663
REFERENCE Z15420
#authors Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft,
I.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
#submission submitted to the Protein Sequence Database, February 1999
#accession T05663
#molecule_type DNA
#residues 1-1055 #label BEV
#cross-references EMBL:AL035539
##experimental_source cultivar Columbia; BAC clone F22I13
GENETICS
#map_position 4
#introns 24/3; 60/3; 99/3; 150/1; 193/1; 216/3; 278/3; 297/2; 336/3;
364/3; 396/3; 419/3; 511/3; 549/3; 577/3; 620/3; 687/1;
735/1; 747/3; 775/3; 825/3; 874/1; 895/3; 927/1; 958/3;
981/2; 1003/3; 1026/3
F22I13.120
#note
#length 1055 #molecular-weight 116148 #checksum 6862
SUMMARY
Query Match 82.4%; Score 42; DB 2; Length 1055;
Best Local Similarity 85.7%; Pred. No. 1.42e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 369 TVDPKS 375
QY 264 SVDPKS 270
|||||:

```

```

RESULT 11
ENTRY D36793 #type complete
TITLE hypothetical protein 67 - ictalurid herpesvirus 1 (strain
auburn 1)
ORGANISM #formal_name ictalurid herpesvirus 1
#note host ictalurus punctatus (channel catfish)
17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change
26-Aug-1999
D36793
#accessions
#authors Davison, A.J.
#submission submitted to GenBank, January 1992
#description Channel catfish virus: a new type of herpesvirus.
#accession D36793
#molecule_type DNA
#residues 1-1556 #label DAV
#cross-references GB:M75136; NID:g331209; PID:g331277
A39447
#authors Davison, A.J.
#journal Virology (1992) 186:9-14
#title Channel catfish virus: a new type of herpesvirus.
#cross-references MUID:92087490
#contents annotation
#note neither protein nor nucleic acid sequence is given
GENETICS
#gene 67
CLASSIFICATION #superfamily ictalurid herpesvirus 1 hypothetical protein 67
SUMMARY #length 1556 #molecular-weight 173577 #checksum 8743
Query Match 82.4%; Score 42; DB 2; Length 1556;
Best Local Similarity 62.5%; Pred. No. 1.42e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 53 SVPEPNSV 60
QY 264 SVDPKSI 271
|||||:

```

```

RESULT 12
ENTRY S53510 #type complete
TITLE pancreatic colipase - pig
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change
17-Mar-1999
ACCESSIONS S53510
REFERENCE S53510
#authors Rugani, N.; Carriere, F.; Thim, L.; Borgstrom, B.; Sarda, L.
#journal Biochim. Biophys. Acta (1995) 1247:185-194
#title Lipid binding and activating properties of porcine pancreatic
colipase split at the Ile(79)-Thr(80) bond.
#cross-references MUID:95210311
#accession S53510
#status preliminary
#molecule_type protein
#residues 1-95 #label RUG
CLASSIFICATION #superfamily colipase
SUMMARY #length 95 #molecular-weight 10309 #checksum 1054
Query Match 80.4%; Score 41; DB 2; Length 95;
Best Local Similarity 71.4%; Pred. No. 2.41e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 1 VPDPRGI 7
QY 265 VPDPSI 271
|||||:
RESULT 13
ENTRY XLP62 #type complete
TITLE colipase II precursor - pig
ALTERNATE_NAMES procolipase II
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE 24-Apr-1984 #sequence_revision 04-Dec-1986 #text_change

```

```
ACCESSIONS 26-Apr-1996
REFERENCE A03162; A90593
#authors Sternby, B.; Engstrom, A.; Hellman, U.; Vihert, A.M.;
#journal Sternby, B.N.H.; Borgstrom, B.
#title Biochim. Biophys. Acta (1984) 784:75-80
#cross-references MUID:84104937
#accession A03162
#molecule_type protein
#residues 1-95 ##label STE
REFERENCE A90593
#authors Charles, M.; Erlanson, C.; Bianchetto, J.; Joffre, J.;
#journal Guidoni, A.; Roversy, M.
#title Biochim. Biophys. Acta (1974) 359:186-197
#cross-references MUID:74290109
#accession A90593
#molecule_type protein
#residues 6-91 ##label CHA
REFERENCE A90594
#authors Erlanson, C.; Charles, M.; Astier, M.; Desnuelle, P.
#journal Biochim. Biophys. Acta (1974) 359:198-203
#title The primary structure of porcine colipase II. II. The
disulfide bridges.
#cross-references MUID:74290110
#contents annotation: disulfide bonds
COMMENT Colipase, a cofactor of triacylglycerol lipase (EC 3.1.1.3), forms
a 1:1 stoichiometric complex with it, enabling it to hydrolyze
its substrate at the lipid-water interface. Without colipase the
enzyme is washed off by bile salts, which are known to have an
inhibitory effect on the lipase.
Residues 6-9 and Arg-92 are considered essential for the function
of colipase in the presence of phospholipids.
CLASSIFICATION #superfamily colipase
KEYWORDS Lipid digestion; lipid hydrolysis; pancreas
FEATURE 1-5
#domain propeptide #status experimental #label PRO\
#product colipase II #status experimental #label MAT\
17-87,23-39,27-63, #disulfide_bonds #status experimental\
28-61,49-69 #binding_site micellar substrate (Phe, Tyr, Tyr, Tyr)
52,55,58,59 #status predicted
SUMMARY #length 95 #molecular-weight 10335 #checksum 795
Query Match 80.4%; Score 41; DB 1; Length 95;
Best Local Similarity 71.1%; Pred. No. 2.41e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 1 VPDPRGI 7
QY 265 VDPDKSI 271
RESULT 14
ENTRY #type complete
TITLE peroxidase (EC 1.11.1.7) - upland cotton
ORGANISM #formal_name Gossypium hirsutum #common_name upland cotton
DATE 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change
26-Aug-1999
ACCESSIONS T10790
REFERENCE #type complete
#authors Ritter, D.; Allen, R.D.; Trolinder, N.; Hughes, D.W.; Galau,
G.A.
#journal Plant Physiol. (1993) 102:1351
#title Cotton cotyledon cDNA encoding a peroxidase.
#cross-references MUID:94105352
#accession T10790
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-332 ##label RIT
#cross-references EMBL:L08199; NID:g157366; PID:g167367

##experimental_source strain Coker 201
CLASSIFICATION #superfamily peroxidase
KEYWORDS oxidoreductase
SUMMARY #length 332 #molecular-weight 37645 #checksum 7211
Query Match 80.4%; Score 41; DB 2; Length 332;
Best Local Similarity 57.1%; Pred. No. 2.41e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 237 IPDPKAV 243
QY 265 VDPDKSI 271
RESULT 15
ENTRY #type complete
TITLE Ca2+-transporting ATPase (EC 3.6.1.38) SERCALa, sarcoplasmic
reticulum - rat
ALTERNATE_NAMES calcium transporting ATPase
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change
22-Jun-1999
ACCESSIONS A48849; I51892
REFERENCE A48849
#authors Wu, K.D.; Lytton, J.
#journal Am. J. Physiol. (1993) 264:C333-C341
#title Molecular cloning and quantification of sarcoplasmic
reticulum Ca(2+)-ATPase isoforms in rat muscles.
#cross-references MUID:93191041
#accession A48849
#molecule_type mRNA
#residues 1-994 #label WU1
#cross-references GB:M99233; NID:g203644; PID:AAA40991.1; PID:g203645
#experimental_source diaphragm
#note sequence extracted from NCBI backbone (NCBIP:126512)
GENETICS
#gene SERCALa
CLASSIFICATION #superfamily Na+/K+-transporting ATPase alpha chain; ATPase
nucleotide-binding domain homology
KEYWORDS ATP; hydrolase
FEATURE 595-768
#domain ATPase nucleotide-binding domain homology #label
ATN
SUMMARY #length 994 #molecular-weight 109408 #checksum 7836
Query Match 80.4%; Score 41; DB 2; Length 994;
Best Local Similarity 50.0%; Pred. No. 2.41e+01;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 193 PVPDPRAV 200
QY 264 SVPDKSI 271
Search completed: Wed May 10 13:48:34 2000
Job time : 7 secs.
```

W P S R E H (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 13:42:10 2000; MasPar time 91.31 Seconds
Tabular output not generated. 2.668 Million cell updates/sec

Title: >US-09-376-430-2
Description: (264-271) from US09376430A.pep (20 of 25)
Perfect Score: 51
Sequence: 1 SVDPDKSI 8

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 20.666; Variance 19.376; scale 1.067

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	46	90.2	558	1	C4BP_RAT	4.59e-01
2	46	90.2	779	1	LEU2_YEAST	4.59e-01
3	43	84.3	112	1	COLCANFA	2.85e+00
4	42	82.4	1556	1	COLCANFA	2.85e+00
5	41	80.4	95	1	COL2_PIG	9.18e+00
6	41	80.4	106	1	COL2_PIG	9.18e+00
7	41	80.4	108	1	COL2_PIG	9.18e+00
8	41	80.4	994	1	ATCB_CHICK	9.18e+00
9	41	80.4	994	1	ATCB_CHICK	9.18e+00
10	41	80.4	997	1	ATCD_HUMAN	9.18e+00
11	41	80.4	997	1	ATCD_HUMAN	9.18e+00
12	41	80.4	997	1	ATCD_RAT	9.18e+00
13	41	80.4	997	1	ATCD_RAT	9.18e+00
14	41	80.4	997	1	ATCD_PIG	9.18e+00
15	41	80.4	1001	1	ATCD_PIG	9.18e+00
16	41	80.4	1001	1	ATCD_PIG	9.18e+00
17	41	80.4	1042	1	ATCE_RABIT	9.18e+00
18	41	80.4	1042	1	ATCE_RABIT	9.18e+00
19	41	80.4	1042	1	ATCE_PIG	9.18e+00
20	41	80.4	1042	1	ATCE_PIG	9.18e+00
21	41	80.4	1043	1	ATCE_HUMAN	9.18e+00
22	40	78.4	1416	1	YNB1_CAEEL	9.18e+00
23	40	78.4	298	1	YNT6_YEAST	1.62e+01

24	40	78.4	590	1	GRK5_HUMAN	G PROTEIN-COUPLED RECE	1.62e+01
25	40	78.4	590	1	GRK5_BOVIN	G PROTEIN-COUPLED RECE	1.62e+01
26	40	78.4	590	1	GRK5_RAT	G PROTEIN-COUPLED RECE	1.62e+01
27	40	78.4	882	1	CAD1_HUMAN	EPITHELIAL-CADHERIN PR	1.62e+01
28	40	78.4	999	1	ATCF_HUMAN	CALCIUM-TRANSPORTING A	1.62e+01
29	40	78.4	999	1	ATCF_RAT	CALCIUM-TRANSPORTING A	1.62e+01
30	40	78.4	1002	1	ATCB_DROME	CALCIUM-TRANSPORTING A	1.62e+01
31	39	76.5	106	1	RS17_METH	30S RIBOSOMAL PROTEIN	2.82e+01
32	39	76.5	109	1	COXA_BOVIN	CYTOCHROME C OXIDASE P	2.82e+01
33	39	76.5	145	1	COXA_MOUSE	CYTOCHROME C OXIDASE P	2.82e+01
34	39	76.5	146	1	COXA_RAT	CYTOCHROME C OXIDASE P	2.82e+01
35	39	76.5	150	1	COXA_HUMAN	CYTOCHROME C OXIDASE P	2.82e+01
36	39	76.5	174	1	RL20_YEAST	60S RIBOSOMAL PROTEIN	2.82e+01
37	39	76.5	308	1	SPE1_DATST	SPERMIDINE SYNTHASE 1	2.82e+01
38	39	76.5	308	1	SPE2_HYONI	SPERMIDINE SYNTHASE 2	2.82e+01
39	39	76.5	314	1	SPEE_NICSY	SPERMIDINE SYNTHASE (E	2.82e+01
40	39	76.5	375	1	PNMT_TOBAC	PUTRESCINE N-METHYLTRA	2.82e+01
41	39	76.5	391	1	DCUP_TOBAC	UROPOPHYRINOGEN DECAR	2.82e+01
42	39	76.5	773	1	LEU2_USTMA	3-ISOPROPYLMALATE DEHY	2.82e+01
43	39	76.5	1736	1	ZOI_HUMAN	TIGHT JUNCTION PROTEIN	2.82e+01
44	39	76.5	1967	1	YG5Q_YEAST	PUTATIVE HELICASE YGR2	2.82e+01
45	39	76.5	3712	1	LMA_DROME	LAMININ ALPHA CHAIN PR	2.82e+01

ALIGNMENTS

RESULT 1	STANDARD;	PRT;	558 AA.
ID C4BP_RAT			
AC Q63514;			
DT 01-NOV-1997 (Rel. 35, Created)			
DT 01-NOV-1997 (Rel. 35, Last sequence update)			
DT 15-JUL-1998 (Rel. 36, Last annotation update)			
DE C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR.			
GN C4BPA.			
OS Rattus norvegicus (Rat).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=LIVER;			
RX MEDLINE; 97166082.			
RA Hillarp A., Wiklund H., Thern A., Dahlback B.;			
RT "Molecular cloning of rat C4b binding protein alpha- and beta-chains:			
RT structural and functional relationships among human, bovine, rabbit,			
RT mouse, and rat proteins.";			
RL J. Immunol. 158:1315-1323(1997).			
CC -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT			
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR			
CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT			
CC ALSO ACCELERATES THE DEGRADATION OF THE C4B/C2A COMPLEX (C3			
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA			
CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S			
CC AND WITH SERUM AMYLOID P COMPONENT.			
CC -!- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.			
CC -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.			
CC -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.			
CC -----			
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC the European Bioinformatics Institute. There are no restrictions on its			
CC use by non-profit institutions as long as its content is in no way			
CC modified and this statement is not removed. Usage by and for commercial			
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC or send an email to: license@isb-sib.ch)			
CC -----			
CC EMBL; Z50051; CAA90391.1;			
DR HSPSP; P10998; IVC.			
DR PFAM; PF00084; sush1; 8.			
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.			
FT SIGNAL 1 13			
FT BY SIMILARITY			
FT C4B-BINDING PROTEIN ALPHA CHAIN.			
FT CHAIN 14 558			
FT 8 X SUSHI (SCR) REPEATS.			
FT DOMAIN 14 502			
FT SUSHI 1.			
FT REPEAT 14 73			

```

FT REPEAT 76 135 SUSHI 2.
FT REPEAT 138 200 SUSHI 3.
FT REPEAT 203 259 SUSHI 4.
FT REPEAT 262 325 SUSHI 5.
FT REPEAT 328 387 SUSHI 6.
FT REPEAT 389 444 SUSHI 7.
FT REPEAT 446 502 SUSHI 8.
FT DISULFID 15 60 BY SIMILARITY.
FT DISULFID 45 72 BY SIMILARITY.
FT DISULFID 77 118 BY SIMILARITY.
FT DISULFID 104 134 BY SIMILARITY.
FT DISULFID 139 182 BY SIMILARITY.
FT DISULFID 168 199 BY SIMILARITY.
FT DISULFID 204 246 BY SIMILARITY.
FT DISULFID 232 258 BY SIMILARITY.
FT DISULFID 263 312 BY SIMILARITY.
FT DISULFID 296 324 BY SIMILARITY.
FT DISULFID 328 351 BY SIMILARITY.
FT DISULFID 329 373 BY SIMILARITY.
FT DISULFID 363 386 BY SIMILARITY.
FT DISULFID 390 431 BY SIMILARITY.
FT DISULFID 417 443 BY SIMILARITY.
FT DISULFID 447 488 BY SIMILARITY.
FT DISULFID 474 501 BY SIMILARITY.
FT DISULFID 509 509 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT DISULFID 521 521 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT CARBOHYD 31 31 POTENTIAL.
FT CARBOHYD 177 177 POTENTIAL.
FT CARBOHYD 186 186 POTENTIAL.
FT CARBOHYD 469 469 POTENTIAL.
FT CARBOHYD 491 491 POTENTIAL.
SQ SEQUENCE 558 AA; 62266 MW; 592FG667ED1B5FF CRC64;

Query Match 90.2%; Score 46; DB 1; Length 558;
Best Local Similarity 75.0%; Pred. No. 4.59e-01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 330 SVPDPKSV 337
Qy 264 SVPDPKSI 271

RESULT 2
ID LEU2_YEAST STANDARD; PRT; 779 AA.
AC P07264;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE 3-ISOPROPYLMALATE DEHYDRATASE (EC 4.2.1.33) (ISOPROPYLMALATE
ISOMERASE) (ALPHA-IPM ISOMERASE) (IPMT).
GN LEU1 OR YGL009C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL125-2B;
RX MEDLINE; 91353082.
RA Skala J., Capieaux E., Balzi E., Chen W., Goffeau A.;
RT "Complete sequence of the Saccharomyces cerevisiae LEU1 gene encoding
isopropylmalate isomerase.";
RL Yeast 7:281-285(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IL125-2B;
RX MEDLINE; 84162042.
RA Hsu Y.-P., Schimmel P.R.;
RT "Yeast LEU1. Repression of mRNA levels by leucine and relationship of
5'-noncoding region to that of LEU2.";
RL J. Biol. Chem. 259:3714-3719(1984).

[4]
RN SEQUENCE OF 740-779 FROM N.A.
RP STRAIN=IL125-2B;
RX MEDLINE; 91353083.
RA Chen W., Balzi E., Capieaux E., Choder M., Goffeau A.;
RT "The DNA sequencing of the 17 kb HindIII fragment spanning the LEU1
and ATE1 loci on chromosome VII from Saccharomyces cerevisiae reveals
the PDR6 gene, a new member of the genetic network controlling
pleiotropic drug resistance.";
RL Yeast 7:287-299(1991).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE -> 2-ISOPROPYLMALATE +
H(2)O (ALSO CATALYSES 2-ISOPROPYLMALATE + H(2)O -> 3-HIDROXY-
4-METHYL-3-CARBOXY-PENTANONE).
CC -1- PATHWAY: SECOND STEP IN LEUCINE BIOSYNTHESIS.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO THE ACNITASE/IPM ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S57886; AAB19612.1;
CC EMBL; Z72531; CAA96709.1;
CC EMBL; K01969; AAA34742.1;
CC EMBL; S58126; CAB31643.1;
CC PIR; A24105; A24105.
CC PIR; S15039; S15039.
CC PIR; S31555; S31555.
CC SGD; L0000942; LEU1.
CC PRINTS; PR00415; ACNITASE.
CC PROSITE; PS00450; ACNITASE_1; 1.
CC PROSITE; PS01244; ACNITASE_2; 1.
CC PFAM; PF00330; acnitate; 1.
CC PFAM; PF00694; Aconitase_C; 1.
KW Leucine biosynthesis; Lyase;
FT METAL 360 360 IRON-SULFUR (4FE-4S).
FT METAL 421 421 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 424 424 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT CONFLICT 291 291 N -> TLKH (IN REF. 1).
FT CONFLICT 423 423 I -> M (IN REF. 1).
FT CONFLICT 459 459 M -> I (IN REF. 1).
FT CONFLICT 744 744 R -> K (IN REF. 1 AND 4).
SQ SEQUENCE 779 AA; 85794 MW; BD409A9702AE3E57 CRC64;

Query Match 90.2%; Score 46; DB 1; Length 779;
Best Local Similarity 75.0%; Pred. No. 4.59e-01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 316 SVPDPKSV 323
Qy 264 SVPDPKSI 271

RESULT 3
ID COL CANFA STANDARD; PRT; 112 AA.
AC P19090;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE COLIPASE PRECURSOR.
GN CLPS.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS.
RX MEDLINE; 91016846.
RA Fukuoka S.-I., Taniguchi Y., Kitagawa Y., Scheele G.;

```


RT "Lipase activation by nonionic detergents. The crystal structure of
 RT the porcine lipase-colipase-tetraethylene glycol monooctyl ether
 RL complex.";
 RL J. Biol. Chem. 271:18007-18016(1996).
 RN [7]
 RP STRUCTURE BY NMR, AND REVISION TO 37.
 RX MEDLINE; 95172049.
 RA Breg J.N., Sarda L., Cozzone P.J., Rugani N., Boelens R., Kapteln R.;
 RT "Solution structure of porcine pancreatic procolipase as determined
 RT from 1H homonuclear two-dimensional and three-dimensional NMR.";
 RL Eur. J. Biochem. 227:663-672(1995).
 CC -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
 CC THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
 CC COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
 CC INHIBITORY EFFECT ON THE LIPASE.
 CC -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
 CC SIGNAL.
 CC -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
 DR PIR; A03162; XLPG2.
 DR PDB; IPCN; 20-DEC-94.
 DR PDB; IPCO; 20-DEC-94.
 DR PDB; ILPA; 01-NOV-94.
 DR PDB; ILPB; 01-NOV-94.
 DR PDB; LEH; 07-DEC-96.
 DR PRINTS; PR00128; COLIPASE.
 DR PROSITE; PS00121; COLIPASE; 1.
 DR PFAM; PF01114; Colipase; 1.
 KW Lipid degradation; Digestion; Pancreas; 3D-structure.
 FT PROPEP 1 5 ENTEROSTATIN, ACTIVATION PEPTIDE.
 FT CHAIN 6 95 COLIPASE II.
 FT SITE 6 9 ESSENTIAL FOR THE FUNCTION OF CLIP IN THE
 FT PRESENCE OF PHOSPHOLIPIDS.
 FT ACT_SITE 92 92 ESSENTIAL FOR THE FUNCTION OF CLIP IN THE
 FT PRESENCE OF PHOSPHOLIPIDS.
 FT DISULFID 17 28
 FT DISULFID 23 39
 FT DISULFID 27 61
 FT DISULFID 49 69
 FT DISULFID 63 87
 FT CONFLICT 37 37 MISSING (IN REF. 2).
 FT CONFLICT 50 50 MISSING (IN REF. 2).
 FT SEQUENCE 95 AA; 10309 MW; 423E04549E196027 CRC64;
 Query Match 80.4%; Score 41; DB 1; Length 95;
 Best Local Similarity 71.4%; Pred. No. 9.18e+00;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 1 VPDPRGI 7
 QY 265 VPDPKSI 271
 RESULT 6
 ID COLA_HORSE STANDARD; PRT; 106 AA.
 AC P02704;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE PROCOLIPASE A PRECURSOR (FRAGMENT).
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Perissodactyla; Equidae; Equus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PANCREAS;
 RX MEDLINE; 94325330.
 RA Cronon I., Granon S., Chapus C., Kerfelec B.;
 RT "Molecular cloning and expression of two horse pancreatic cDNA
 RT encoding colipase A and B.";
 RL Biochim. Biophys. Acta 1213:357-360(1994).
 RN [2]
 RP SEQUENCE OF 12-106.
 RX MEDLINE; 84104937.

RA Sternby B., Engstrom A., Hellman U., Vihert A.M., Sternby N.H.,
 RA Borgstrom B.;
 RT "The primary sequence of human pancreatic colipase.";
 RL Biochim. Biophys. Acta 784:75-80(1984).
 RN [3]
 RP SEQUENCE OF 12-106.
 RX MEDLINE; 82186702.
 RA Pierrot M., Astier J.-P., Astier M., Charles M., Drenth J.;
 RT "Pancreatic colipase: crystallographic and biochemical aspects.";
 RL Eur. J. Biochem. 123:347-354(1982).
 RN [4]
 RP SEQUENCE OF 12-66.
 RX MEDLINE; 81021166.
 RA Julien R., Bechis G., Gregoire J., Rathelot J., Rochat H., Sarda L.;
 RT "Evidence for the existence of two isocollipases in horse pancreas.";
 RL Biochem. Biophys. Res. Commun. 95:1245-1252(1980).
 CC -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
 CC THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
 CC COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
 CC INHIBITORY EFFECT ON THE LIPASE.
 CC -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
 CC SIGNAL.
 CC -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X74503; CAA52611.1; -.
 DR PIR; A03164; XLHOA.
 DR HSSP; P02703; LEH.
 DR PROSITE; PS00121; COLIPASE; 1.
 DR PFAM; PF01114; Colipase; 1.
 KW Lipid degradation; Digestion; Pancreas; Signal.
 FT NON_TER 1 1
 FT SIGNAL <1 11 ENTEROSTATIN, ACTIVATION PEPTIDE.
 FT PROPEP 12 16 COLIPASE A.
 FT CHAIN 17 106 BY SIMILARITY.
 FT DISULFID 28 39 BY SIMILARITY.
 FT DISULFID 34 50 BY SIMILARITY.
 FT DISULFID 38 72 BY SIMILARITY.
 FT DISULFID 60 80 BY SIMILARITY.
 FT DISULFID 74 98 BY SIMILARITY.
 FT BINDING 63 63 BILE SALT MICELLES.
 FT CONFLICT 33 33 Q -> E (IN REF. 2 AND 3).
 FT CONFLICT 43 43 S -> E (IN REF. 2, 3 AND 4).
 FT CONFLICT 100 100 D -> N (IN REF. 3).
 FT CONFLICT 103 103 R -> K (IN REF. 3).
 FT CONFLICT 106 106 E -> ER (IN REF. 3).
 FT SEQUENCE 106 AA; 11388 MW; CDDC36747D94810C CRC64;
 Query Match 80.4%; Score 41; DB 1; Length 106;
 Best Local Similarity 50.0%; Pred. No. 9.18e+00;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 11 AVDPGRGV 18
 QY 264 SVPDPKSI 271
 RESULT 7
 ID COLB_HORSE STANDARD; PRT; 108 AA.
 AC P02705;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE PROCOLIPASE B PRECURSOR (FRAGMENT).
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Perissodactyla; Equidae; Equus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PANCREAS;
 RX MEDLINE; 94325330.
 RA Crenon I., Granon S., Chapus C., Kerfelec B.;
 RT "Molecular cloning and expression of two horse pancreatic cDNA
 RL encoding colipase A and B";
 RN Biochim. Biophys. Acta 1213:357-360(1994).
 [2]
 RN SEQUENCE OF 14-108.
 RP MEDLINE; 82046794.
 RX Bonicel J.J., Couchoud P.M., Foglizzo E., Desnuelle P., Chapus C.;
 RA "Amino acid sequence of horse colipase B";
 RL Biochim. Biophys. Acta 669:39-45(1981).
 [3]
 RN SEQUENCE OF 14-68.
 RP MEDLINE; 81021166.
 RX Julien R., Bechis G., Gregoire J., Rathelot J., Rochat H., Sarda L.;
 RA "Evidence for the existence of two isocollipases in horse pancreas";
 RL Biochim. Biophys. Res. Commun. 95:1245-1252(1980).
 CC [1]- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
 CC THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
 CC COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
 CC INHIBITORY EFFECT ON THE LIPASE.
 CC [1]- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
 CC SIGNAL.
 CC [1]- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X74344; CAA52391.1; -.
 DR PIR; A03165; XIHOB.
 DR HSP; P02703; IETH.
 DR PROSITE; PS00121; COLIPASE; 1.
 DR PFAM; PF01114; Colipase; 1.
 DR KW Lipid degradation; Digestion; Pancreas; Signal.
 FT NON_TER 1
 FT SIGNAL <1 13
 FT PROPEP 14 18
 FT CHAIN 19 108
 FT BINDING 65 65
 FT DISULFID 30 41
 FT BY SIMILARITY.
 FT DISULFID 36 52
 FT BY SIMILARITY.
 FT DISULFID 40 74
 FT BY SIMILARITY.
 FT DISULFID 62 82
 FT BY SIMILARITY.
 FT DISULFID 76 100
 FT BY SIMILARITY.
 FT CONFLICT 35 35
 FT Q -> E (IN REF. 3).
 FT CONFLICT 42 42
 FT H -> T (IN REF. 3).
 FT CONFLICT 108 108
 FT E -> ER (IN REF. 2).
 SQ SEQUENCE 108 AA; 11618 MW; 37A7F49ABAA08B30 CRC64;
 Query Match 80.4%; Score 41; DB 1; Length 108;
 Best Local Similarity 50.0%; Pred. No. 9.18e+00;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 13 AVDPDRGV 20
 QY 264 SVPDPKSI 271
 RESULT 8
 ID ATCB CHICK STANDARD; PRT; 994 AA.
 AC P13585;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)

DE DE CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC RETICULUM TYPE, FAST TWITCH
 OS SKELETAL MUSCLE, ADULT ISOFORM (EC 3.6.1.38) (CALCIUM PUMP) (SERCA1).
 OC Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
 OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 89313743.
 RA Karin N.J., Kaprielian Z., Fambrough D.M.;
 RT "Expression of avian Ca2+-ATPase in cultured mouse myogenic cells";
 RL Mol. Cell. Biol. 9:1978-1986(1989).
 RN [2]
 RP REVISIONS.
 RA Karin N.J.;
 RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.
 CC [1]- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE
 CC HYDROLYSIS OF ATP COUPLED WITH THE TRANSPORT OF THE CALCIUM.
 CC [1]- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
 CC [1]- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. SARCOPLASMIC
 CC RETICULUM.
 CC [1]- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M26064; AAA48609.1; -.
 DR PIR; A32792; A32792.
 DR PRINTS; PR00119; CATATPASE.
 DR PROSITE; PS00154; ATPASE_E1_E2; 1.
 DR PFAM; PF00122; E1-E2_ATPase; 1.
 DR KW Hydrolyase; Calcium transport; Transmembrane; Phosphorylation;
 KW Magnesium; ATP-binding; Alternative splicing.
 FT DOMAIN 1 59
 FT TRANSSEM 60 78
 FT POTENTIAL.
 FT DOMAIN 79 86
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 87 107
 FT POTENTIAL.
 FT DOMAIN 108 259
 FT CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 260 279
 FT POTENTIAL.
 FT DOMAIN 280 297
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 298 315
 FT POTENTIAL.
 FT DOMAIN 316 771
 FT CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 772 795
 FT POTENTIAL.
 FT DOMAIN 796 836
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 837 859
 FT POTENTIAL.
 FT DOMAIN 860 894
 FT CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 895 913
 FT POTENTIAL.
 FT DOMAIN 914 930
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 931 950
 FT POTENTIAL.
 FT DOMAIN 951 994
 FT CYTOPLASMIC (POTENTIAL).
 FT MOD_RES 351 351
 FT PHOSPHORYLATION.
 FT BINDING 515 515
 FT ATP (BY SIMILARITY).
 SQ SEQUENCE 994 AA; 109023 MW; 1F490D32F3EC319A CRC64;
 Query Match 80.4%; Score 41; DB 1; Length 994;
 Best Local Similarity 50.0%; Pred. No. 9.18e+00;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 193 PVPDPRAV 200
 QY 264 SVPDPKSI 271
 RESULT 9
 ID ATCB RABIT STANDARD; PRT; 994 AA.
 AC P11719;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC RETICULUM TYPE, FAST TWITCH
 GN SKELETAL MUSCLE, ADULT ISOFORM (EC 3.6.1.38) (CALCIUM PUMP) (SERCAL).
 GN ATP2A1.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 [1]
 RN SEQUENCE OF 1-993 FROM N.A.
 RP MEDLINE; 86133552.
 RA Brandl C.J., Green N.M., Korczak B., McLennan D.H.;
 RX "Two Ca2+ ATPase genes: homologies and mechanistic implications of
 RT deduced amino acid sequences.";
 RL J. Biol. Chem. 263:3768-3774(1988).
 [2]
 RN SEQUENCE OF 973-994 FROM N.A.
 RP MEDLINE; 87137681.
 RA Brandl C.J., Deleon S., Martin D.R., McLennan D.H.;
 RX "Adult forms of the Ca2+ATPase of sarcoplasmic reticulum. Expression
 RT in developing skeletal muscle.";
 RL J. Biol. Chem. 263:3768-3774(1988).
 CC -!- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE
 CC HYDROLYSIS OF ATP COUPLED WITH THE TRANSPORT OF THE CALCIUM.
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O -> ADP + ORTHOPHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. SARCOPLASMIC
 CC RETICULUM.
 CC -!- ALTERNATIVE PRODUCTS: FAST TWITCH SKELETAL MUSCLE, NEONATAL AND
 CC ADULT ISOFORM ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- DISEASE: DEFICIENCY OF THE ENZYME IN FAST TWITCH MUSCLE CAUSES
 CC BRODY'S DISEASE, A DISORDER OF SKELETAL MUSCLE RELAXATION.
 CC -!- MISCELLANEOUS: IN RABBIT, HOMOLOGOUS ENZYMES ARE FOUND IN FAST
 CC TWITCH AND SLOW TWITCH SKELETAL MUSCLE SARCOPLASMIC RETICULUM.
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; M12898; ; NOT ANNOTATED_CDS.
 DR EMBL; M15358; AAA31167.1; .
 DR PRINTS; PR00119; CATATPASE.
 DR PROSITE; PS00154; ATPASE_E1_E2; 1.
 DR PFAM; PF00122; E1-E2 ATPase; 1.
 DR KWL Hydrolyase; Calcium transport; Transmembrane; Phosphorylation;
 KW Magnesium; ATP-binding; Alternative splicing.
 FT DOMAIN 1 59
 FT TRANSMEM 60 78
 FT POTENTIAL.
 FT DOMAIN 79 86
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 87 107
 FT POTENTIAL.
 FT DOMAIN 108 259
 FT CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 260 279
 FT POTENTIAL.
 FT DOMAIN 280 297
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 298 315
 FT POTENTIAL.
 FT DOMAIN 316 771
 FT CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 772 795
 FT POTENTIAL.
 FT DOMAIN 796 836
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 837 859
 FT POTENTIAL.
 FT DOMAIN 860 894
 FT CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 895 913
 FT POTENTIAL.
 FT DOMAIN 914 930
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 931 950
 FT POTENTIAL.
 FT DOMAIN 951 994
 FT CYTOPLASMIC (POTENTIAL).
 FT MOD_RES 351 351
 FT PHOSPHORYLATION.
 FT BINDING 515 515
 FT ATP (BY SIMILARITY).
 SQ SEQUENCE 994 AA; 109489 MW; 148C462A8653A3E5 CRC64;

Query Match 80.4%; Score 41; DB 1; Length 994;
 Best Local Similarity 50.0%; Pred. No. 9.18e+00;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 193 PVPDPRAV 200
 QY 264 SVPDPKSI 271
 RESULT 10
 ID ATCD_HUMAN STANDARD; PRT; 997 AA.
 AC P16614;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC RETICULUM TYPE, SLOW TWITCH
 DE SKELETAL MUSCLE ISOFORM (EC 3.6.1.38) (ENDOPLASMIC RETICULUM CLASS 1
 DE CA(2+) ATPASE) (CALCIUM PUMP) (SERCA2) (HK2).
 GN ATP2A2 OR ATP2B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
 [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE-KIDNEY;
 RX MEDLINE; 89008384.
 RA Lytton J., MacLennan D.H.;
 RT "Molecular cloning of cDNAs from human kidney coding for two
 RT alternatively spliced products of the cardiac Ca2+-ATPase gene.";
 RL J. Biol. Chem. 263:15024-15031(1988).
 CC -!- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE
 CC HYDROLYSIS OF ATP COUPLED WITH THE TRANSPORT OF THE CALCIUM.
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O -> ADP + ORTHOPHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. SARCOPLASMIC
 CC AND ENDOPLASMIC RETICULUM.
 CC -!- ALTERNATIVE PRODUCTS: SLOW TWITCH SR/CLASS I ER AND CLASS 2 ER
 CC CA ATPASES ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; M23115; AAA53194.1; .
 DR EMBL; M23278; AAA52758.1; .
 DR EMBL; M23116; AAA52758.1; JOINED.
 DR FIR; B31981; B31981.
 DR HSC-2DPAGE; P16614; HUMAN.
 DR MIM; 108740; .
 DR PRINTS; PR00119; CATATPASE.
 DR PROSITE; PS00154; ATPASE_E1_E2; 1.
 DR PFAM; PF00122; E1-E2 ATPase; 1.
 DR KWL Hydrolyase; Calcium transport; Transmembrane; Phosphorylation;
 KW Magnesium; ATP-binding; Multigene family; Alternative splicing.
 FT DOMAIN 1 59
 FT TRANSMEM 60 78
 FT POTENTIAL.
 FT DOMAIN 79 86
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 87 107
 FT POTENTIAL.
 FT DOMAIN 108 259
 FT CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 260 279
 FT POTENTIAL.
 FT DOMAIN 280 297
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 298 315
 FT POTENTIAL.
 FT DOMAIN 316 771
 FT CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 772 794
 FT POTENTIAL.
 FT DOMAIN 795 835
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 836 858
 FT POTENTIAL.
 FT DOMAIN 859 893
 FT CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 894 912
 FT POTENTIAL.
 FT DOMAIN 913 929
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 930 949
 FT POTENTIAL.

FT DOMAIN 950 997 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 351 351 PHOSPHORYLATION.
FT BINDING 514 514 ATP (BY SIMILARITY).
SQ SEQUENCE 997 AA; 109690 MW; DD57D12A2B24FEC1 CRC64;
Query Match 80.4%; Score 41; DB 1; Length 997;
Best Local Similarity 50.0%; Pred. No. 9.18e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 193 PVPDPRAV 200
Qy 264 SVPDPKSI 271
RESULT 11
ID ATCD.RABIT STANDARD; PRT; 997 AA.
AC P04192; 1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC RETICULUM TYPE, SLOW TWITCH
DE SKELETAL MUSCLE ISOFORM (EC 3.6.1.38) (ENDOPLASMIC RETICULUM CLASS 1)
DE (CALCIUM PUMP) (SERCA2).
GN Oryctolagus cuniculus (Rabbit).
OS Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85296300.
RA McGinnis D.H., Brandl C.J., Korczak B., Green N.M.;
RT "Amino-acid sequence of a Ca2+ + Mg2+-dependent ATPase from rabbit
RT muscle sarcoplasmic reticulum, deduced from its complementary DNA
RT sequence.";
RL Nature 316:696-700(1985).
CC -!- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE
CC HYDROLYSIS OF ATP COUPLED WITH THE TRANSPORT OF THE CALCIUM.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. SARCOPLASMIC
CC AND ENDOPLASMIC RETICULUM.
CC -!- ALTERNATIVE PRODUCTS: SLOW TWITCH SR/CLASS1 ER AND CLASS 2 ER
CC CA ATPASES ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC -!- MISCELLANEOUS: THE SLOW TWITCH ENZYME IS LOW IN ACTIVITY.
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X02814; CAA26583.1;
CC FIRM: A01076; PIRBSC.
CC PRINTS: PS00119; CATATPASE.
CC PROSITE: PS00154; ATPASE_E1_E2; 1.
CC PFAM: PF00122; E1-E2_ATPase; 1.
KW Hydrolase; Calcium transport; Transmembrane; Phosphorylation;
KW Magnesium; ATP-binding; Alternative splicing.
FT DOMAIN 1 59 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 60 78 POTENTIAL.
FT DOMAIN 79 86 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 87 107 POTENTIAL.
FT DOMAIN 108 259 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 260 279 POTENTIAL.
FT DOMAIN 280 297 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 298 315 POTENTIAL.
FT DOMAIN 316 771 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 772 794 POTENTIAL.
FT DOMAIN 795 835 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 836 858 POTENTIAL.

FT DOMAIN 859 912 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 894 929 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 913 929 POTENTIAL.
FT TRANSSEM 930 949 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 950 997 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 351 351 PHOSPHORYLATION.
FT BINDING 514 514 ATP (BY SIMILARITY).
SQ SEQUENCE 997 AA; 109644 MW; 1CE7B88F8A7BC12D CRC64;
Query Match 80.4%; Score 41; DB 1; Length 997;
Best Local Similarity 50.0%; Pred. No. 9.18e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 193 PVPDPRAV 200
Qy 264 SVPDPKSI 271
RESULT 12
ID ATCD.RAT STANDARD; PRT; 997 AA.
AC P11508;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC RETICULUM TYPE, SLOW TWITCH
DE SKELETAL MUSCLE ISOFORM (EC 3.6.1.38) (ENDOPLASMIC RETICULUM CLASS 1)
DE CA(2+) ATPASE) (CALCIUM PUMP) (SERCA2).
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 8908385.
RA Guneski-Hamblin A.M., Greeb J., Shull G.E.;
RT "A novel Ca2+ pump expressed in brain, kidney, and stomach is encoded
RT by an alternative transcript of the slow-twitch muscle sarcoplasmic
RT reticulum Ca-ATPase gene. Identification of cDNAs encoding Ca2+ and
RT other cation-transporting ATPases using an oligonucleotide probe
RT derived from the ATP-binding site.";
RL J. Biol. Chem. 263:15032-15040(1988).
CC [2]
CC SEQUENCE FROM N.A.
CC TISSUE=HEART;
CC MEDLINE: 89252068.
CC Lompre A.M., de la Bastie D., Boheler K.R., Schwartz K.;
CC "Characterization and expression of the rat heart sarcoplasmic
CC reticulum Ca2+-ATPase mRNA.";
CC FEBS Lett. 249:35-41(1989).
CC -!- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE
CC HYDROLYSIS OF ATP COUPLED WITH THE TRANSPORT OF THE CALCIUM.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. SARCOPLASMIC
CC AND ENDOPLASMIC RETICULUM.
CC -!- ALTERNATIVE PRODUCTS: SLOW TWITCH SR/CLASS1 ER AND CLASS 2 ER
CC CA ATPASES ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J04023; AAA40786.1;
CC EMBL: X15635; CAA33645.1;
CC PIR: B31982; B31982.
CC PIR: S04269; S04269.
CC PRINTS: PR00119; CATATPASE.

DR PROSITE: PS00154; ATPASE_E1_E2; 1.
 DR PFAM: PF00122; E1-E2_ATPase; 1.
 KW Hydrolyase; Calcium transport; Transmembrane; Phosphorylation;
 FT DOMAIN 1 59
 FT DOMAIN 60 78
 FT DOMAIN 79 86
 FT DOMAIN 87 107
 FT DOMAIN 108 259
 FT DOMAIN 260 279
 FT DOMAIN 280 297
 FT DOMAIN 298 315
 FT DOMAIN 316 771
 FT DOMAIN 772 794
 FT DOMAIN 795 835
 FT DOMAIN 836 858
 FT DOMAIN 859 893
 FT DOMAIN 894 912
 FT DOMAIN 913 929
 FT DOMAIN 930 949
 FT DOMAIN 950 997
 FT MOD_RES 351 351
 FT BINDING 514 514
 FT CONFLICT 272 272
 FT CONFLICT 288 288
 FT CONFLICT 557 557
 SQ SEQUENCE 997 AA; 109680 MW; 3A14A6B604E97002 CRC64;
 Query Match 80.4%; Score 41; DB 1; Length 997;
 Best Local Similarity 50.0%; Pred. No. 9.18e+00;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 193 PVPDPRAV 200
 Qy 264 SVPDPKSI 271
 RESULT 13
 ID ATCD_PIG STANDARD; PRT; 997 AA.
 AC P11606;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC RETICULUM TYPE, SLOW TWITCH
 DE SKELETAL MUSCLE ISOFORM (EC 3.6.1.38) (ENDOPLASMIC RETICULUM CLASS 1
 DE CA(2+) ATPASE) (CALCIUM PUMP) (SERCA2).
 GN ATP2A2.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=STOMACH SMOOTH MUSCLE;
 RX MEDLINE: 89350878;
 RA Eggermont J.A., Wuytack F., de Jaegere S., Nelles L., Casteels R.;
 RT "Evidence for two isoforms of the endoplasmic-reticulum Ca2+ pump in
 RT pig smooth muscle.";
 RL Biochem. J. 260:757-761(1989).
 RN [2]
 RP SEQUENCE OF 842-997 FROM N.A.
 RX MEDLINE: 91198156.
 RA Eggermont J.A., Wuytack F., Casteels R.;
 RT "Characterization of the 3' end of the pig sarcoplasmic/endoplasmic-
 RT reticulum Ca2+ pump gene 2.";
 RL Biochim. Biophys. Acta 1088:448-451(1991).
 CC -!- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE
 CC HYDROLYSIS OF ATP COUPLED WITH THE TRANSPORT OF THE CALCIUM.
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. SARCOPLASMIC
 CC AND ENDOPLASMIC RETICULUM.
 CC -!- ALTERNATIVE PRODUCTS: SLOW TWITCH SR/CLASS I ER AND CLASS 2 ER
 CC CA ATPASES ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY

CC (E1-E2 ATPASES).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X15073; CAA33169.1; -
 DR EMBL: X53754; CAA37783.1; -
 DR PIR: S04651; S04651.
 DR PRINTS: PR00119; CATATPASE.
 DR PROSITE: PS00154; ATPASE_E1_E2; 1.
 DR PFAM: PF00122; E1-E2_ATPase; 1.
 KW Hydrolyase; Calcium transport; Transmembrane; Phosphorylation;
 KW Magnesium; ATP-binding; Alternative splicing
 FT DOMAIN 1 59
 FT DOMAIN 60 78
 FT DOMAIN 79 86
 FT TRANSMEM 87 107
 FT DOMAIN 108 259
 FT TRANSMEM 260 279
 FT DOMAIN 280 297
 FT TRANSMEM 298 315
 FT DOMAIN 316 771
 FT TRANSMEM 772 794
 FT DOMAIN 795 835
 FT TRANSMEM 836 858
 FT DOMAIN 859 893
 FT TRANSMEM 894 912
 FT DOMAIN 913 929
 FT TRANSMEM 930 949
 FT DOMAIN 950 997
 FT MOD_RES 351 351
 FT BINDING 514 514
 SQ SEQUENCE 997 AA; 109726 MW; 66697619E312CA3E CRC64;
 Query Match 80.4%; Score 41; DB 1; Length 997;
 Best Local Similarity 50.0%; Pred. No. 9.18e+00;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 193 PVPDPRAV 200
 Qy 264 SVPDPKSI 271
 RESULT 14
 ID ATCD_FELCA STANDARD; PRT; 997 AA.
 AC Q00779;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC RETICULUM TYPE, SLOW TWITCH
 DE SKELETAL MUSCLE ISOFORM (EC 3.6.1.38) (ENDOPLASMIC RETICULUM CLASS 1
 DE CA(2+) ATPASE) (CALCIUM PUMP) (SERCA2).
 GN ATP2A2.
 OS Felis silvestris catus (Cat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Carnivora; Fissipedia; Felidae; Felis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HEART;
 RX MEDLINE: 92305061.
 RA Gabel A.M., Gallien T.N., Dantzer-Whitworth T., Bowes J.,
 RA Menick D.R.;
 RT "Sequence of the feline cardiac sarcoplasmic reticulum
 RT Ca(2+)-ATPase.";
 RL Biochim. Biophys. Acta 1131:203-206(1992).
 CC -!- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE
 CC HYDROLYSIS OF ATP COUPLED WITH THE TRANSPORT OF THE CALCIUM.
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.

US-09-376-430-2-20.rsp

Thu May 11 06:50:10 2000

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. SARCOPLASMIC
CC AND ENDOPLASMIC RETICULUM.
CC -!- ALTERNATIVE PRODUCTS: SLOW TWITCH SR/CLASS I ER AND CLASS 2 ER
CC CA ATPASES ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z11500; CAAT7576.1; -.
CC PIR: S23444; S23444.
CC PRINTS: PR00119; CATATPASE.
CC PFAM: PF00122; E1-E2_ATPase; 1.
CC KW Hydrolyase; Calcium transport; Transmembrane; Phosphorylation;
CC Mgnesium; ATP-binding; Alternative splicing.
CC FT DOMAIN 1 59 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 60 78 POTENTIAL.
CC FT DOMAIN 79 86 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 87 107 POTENTIAL.
CC FT DOMAIN 108 259 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 260 279 POTENTIAL.
CC FT DOMAIN 280 297 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 298 315 POTENTIAL.
CC FT DOMAIN 316 771 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 772 794 POTENTIAL.
CC FT DOMAIN 795 835 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 836 858 POTENTIAL.
CC FT DOMAIN 859 893 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 894 912 POTENTIAL.
CC FT DOMAIN 913 929 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 930 949 POTENTIAL.
CC FT DOMAIN 950 997 CYTOPLASMIC (POTENTIAL).
CC FT MOD_RES 351 351 PHOSPHORYLATION (BY SIMILARITY).
CC FT BINDING 514 514 ATP (BY SIMILARITY).
CC SQ SEQUENCE 997 AA; 109712 MW; CEE18D1A1ADA738F CRC64;

Query Match 80.4%; Score 41; DB 1; Length 997;
Best Local Similarity 50.0%; Pred. No. 9.18e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 193 PVPDPRAV 200
QY 264 SVPDPKSI 271

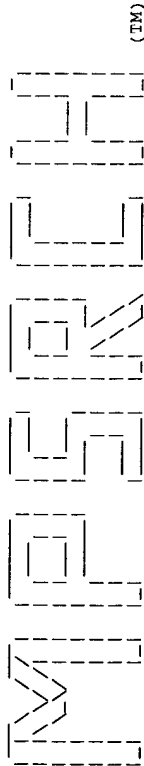
RESULT 15
ID ATCA_RABIT STANDARD; PRT: 1001 AA.
AC P04191;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC RETICULUM TYPE, FAST TWITCH
DE SKELETAL MUSCLE, NEONATAL ISOFORM (EC 3.6.1.38) (CALCIUM PUMP)
DE (SERCAL).
GN ATP2A1
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 86133552.
RA Brandl C.J., Green N.M., Korczak B., McLennan D.H.;
RT Two Ca2+ ATPase genes: homologies and mechanistic implications of
RT deduced amino acid sequences.;
RL Cell 44:597-607(1986).
RN [2]

RP SEQUENCE OF 973-1001 FROM N.A.
RX MEDLINE: 87137681.
RA Brandl C.J., Deleon S., Martin D.R., McLennan D.H.;
RT "Adult forms of the Ca2+-ATPase of sarcoplasmic reticulum. Expression
RT in developing skeletal muscle.";
RL J. Biol. Chem. 262:3768-3774(1987).
CC -!- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE
CC HYDROLYSIS OF ATP COUPLED WITH THE TRANSPORT OF THE CALCIUM.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. SARCOPLASMIC
CC RETICULUM.
CC -!- ALTERNATIVE PRODUCTS: FAST TWITCH SKELETAL MUSCLE, NEONATAL AND
CC ADULT ISOFORM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DISEASE: DEFICIENCY OF THE ENZYME IN FAST TWITCH MUSCLE CAUSES
CC BRODY'S DISEASE, A DISORDER OF SKELETAL MUSCLE RELAXATION.
CC -!- MISCELLANEOUS: IN RABBIT, HOMOLOGOUS ENZYMES ARE FOUND IN FAST
CC TWITCH AND SLOW TWITCH SKELETAL MUSCLE SARCOPLASMIC RETICULUM.
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M12898; AAA31165.1; -.
CC EMBL: M15351; AAA31166.1; -.
CC PIR: A01075; PWRBFC.
CC PRINTS: PR00119; CATATPASE.
CC PFAM: PF00122; E1-E2_ATPase; 1.
CC KW Hydrolyase; Calcium transport; Transmembrane; Phosphorylation;
CC Mgnesium; ATP-binding; Alternative splicing.
CC FT DOMAIN 1 59 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 60 78 POTENTIAL.
CC FT DOMAIN 79 86 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 87 107 POTENTIAL.
CC FT DOMAIN 108 259 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 260 279 POTENTIAL.
CC FT DOMAIN 280 297 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 298 315 POTENTIAL.
CC FT DOMAIN 316 771 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 772 795 POTENTIAL.
CC FT DOMAIN 796 835 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 837 859 POTENTIAL.
CC FT DOMAIN 860 894 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 895 913 POTENTIAL.
CC FT DOMAIN 914 930 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 931 950 POTENTIAL.
CC FT DOMAIN 951 1001 CYTOPLASMIC (POTENTIAL).
CC FT MOD_RES 351 351 PHOSPHORYLATION.
CC FT BINDING 515 515 ATP (BY SIMILARITY).
CC SQ SEQUENCE 1001 AA; 110458 MW; IF0D8C36CF975266 CRC64;

Query Match 80.4%; Score 41; DB 1; Length 1001;
Best Local Similarity 50.0%; Pred. No. 9.18e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 193 PVPDPRAV 200
QY 264 SVPDPKSI 271

Search completed: Wed May 10 13:43:49 2000
Job time : 99 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 13:44:06 2000; MasPar time 233.64 Seconds
Tabular output not generated. 2.374 Million cell updates/sec

Title: >US-09-376-430-2
Description: (264-271) from US09376430A.pep (20 of 25)
Perfect Score: 51
Sequence: 1 SVPDPKSI 8

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organella
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 19.913; Variance 18.290; scale 1.089

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	48	94.1	Q19542	F18C12.1 PROTEIN.	1.31e-01
2	45	88.2	Q43730	PEROXIDASE PRECURSOR (9.41e-01
3	45	88.2	Q42579	PEROXIDASE ATP1A PRECU	9.41e-01
4	44	86.3	Q70083	SARCOPLASMIC/ENDOPLASM	1.78e+00
5	44	86.3	Q70084	SARCOPLASMIC/ENDOPLASM	1.78e+00
6	44	86.3	Q9XU13	K11D9.2B PROTEIN.	1.78e+00
7	44	86.3	Q9XG66	K11D9.2A PROTEIN.	1.78e+00
8	43	84.3	Q62226	F35E2.4 PROTEIN.	3.35e+00
9	43	84.3	Q92258	T15B16.1 PROTEIN.	3.35e+00
10	43	84.3	Q16657	C03H5.3 PROTEIN.	3.35e+00
11	43	84.3	Q64970	CATIONIC PEROXIDASE 2	3.35e+00
12	43	84.3	Q9XET6	PEROXIDASE.	3.35e+00
13	43	84.3	Q81799	HYPOTHETICAL 57.0 KD P	3.35e+00
14	43	84.3	Q95256	INTERLEUKIN-18 RECEPT	3.35e+00
15	43	84.3	Q17315	SARCO/ENDOPLASMIC RETI	3.35e+00
16	43	84.3	Q17314	SARCO/ENDOPLASMIC RETI	3.35e+00
17	42	82.4	Q53805	PRE-PHENOMYCIN PRECURS	6.21e+00
18	42	82.4	Q27307	FORMATE HYDROGENLYASE.	6.21e+00
19	42	82.4	Q9XEL2	SECRETORY PEROXIDASE.	6.21e+00
20	42	82.4	Q9X268	OLIGOPEPTIDE ABC TRANS	6.21e+00

21	42	82.4	1011	5	096527	CALCIUM ATPASE 2.	6.21e+00
22	42	82.4	1865	10	081909	T7123.15 PROTEIN.	6.21e+00
23	41	80.4	332	10	008671	PEROXIDASE PRECURSOR (1.14e+01
24	41	80.4	502	11	054994	SARCOPLASMIC RETICULUM	1.14e+01
25	41	80.4	652	3	075011	N-TERMINAL SERINE/THRE	1.14e+01
26	41	80.4	899	5	044872	R06B10.4 PROTEIN.	1.14e+01
27	41	80.4	994	11	064578	CALCIUM TRANSPORTING A	1.14e+01
28	41	80.4	994	4	014984	CA2+ ATPASE OF FAST-TW	1.14e+01
29	41	80.4	997	13	003669	CA2+ ATPASE.	1.14e+01
30	41	80.4	997	6	045674	SARCOPLASMIC RETICULUM	1.14e+01
31	41	80.4	998	11	055143	SARCO-ENDOPLASMIC RETI	1.14e+01
32	41	80.4	1001	4	014983	CA2+ ATPASE OF FAST-TW	1.14e+01
33	40	78.4	147	11	050252	FORMATE HYDROGENLYASE	2.07e+01
34	40	78.4	185	2	09X5F0	RIBOSOME RECYCLING FAC	2.07e+01
35	40	78.4	229	2	P94406	DNA FOR 25-36 DEGREE R	2.07e+01
36	40	78.4	360	4	Q9Y5P6	GDP-MANNOSE PYROPHOSPH	2.07e+01
37	40	78.4	728	3	Q05812	CHROMOSOME XII COSMID	2.07e+01
38	40	78.4	890	3	060040	100 KDA PROTEIN.	2.07e+01
39	40	78.4	895	10	080973	PUTATIVE MUDR PROTEIN.	2.07e+01
40	40	78.4	942	10	092UQ8	PUTATIVE TRANSPOSON MU	2.07e+01
41	40	78.4	998	4	075501	SARCO-/ENDOPLASMIC RET	2.07e+01
42	40	78.4	999	11	Q64518	SARCO-ENDOPLASMIC RETIC	2.07e+01
43	40	78.4	1029	4	060900	SARCO/ENDOPLASMIC RETI	2.07e+01
44	40	78.4	1038	11	Q64517	SARCO/ENDOPLASMIC RETI	2.07e+01
45	40	78.4	1163	13	Q9YHV2	TIGHT JUNCTION PROTEIN	2.07e+01

ALIGNMENTS

RESULT 1
ID Q19542 PRELIMINARY: PRT: 4131 AA.
AC Q19542;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE F18C12.1 PROTEIN.
GN F18C12.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA HARRIS B.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA SKALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans";
RL Nature 358:32-38(1994).
DR EMBL; Z75536; CAA9830.1; -.
DR HSP; P02633; 1B1G.
SQ SEQUENCE 4131 AA; 459655 MW; A7FF0E8 CRC32;

Query Match 94.1%; Score 48; DB 5; Length 4131;
Best Local Similarity 87.5%; Pred. No. 1.31e-01;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
bb '2157 SVPDPKNI 2164
QY 264 SVPDPKSI 271

```

RESULT 2
ID Q43730 PRELIMINARY; PRT; 330 AA.
AC Q43730;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PEROXIDASE PRECURSOR (EC 1.11.1.7) (MYELOPEROXIDASE).
GN YPR9 OR PRX1.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-CV COLUMBIA;
RC TISSUE-LEAVES, STEMS, FLOWERS, ROOTS, AND SILIQUES;
RA CAPELLI N., TOGNOLLI M., FLACH J., OVERNEY S., PENEL C., GREPPIN H.,
RA SIMON P.;
RL Plant Physiol. 112:446-446(1996).
DR EMBL; X98313; CAA66957.1; -.
DR HSSP; 6564; Arath.Ypr9; 6564.
DR PFAM; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
KW Signal; Peroxidase; Oxidoreductase.
FT SIGNAL 23 POTENTIAL.
FT CHAIN 1 230 POTENTIAL.
FT CHAIN 24 330 POTENTIAL.
SQ SEQUENCE 330 AA; 37295 MW; 88F9CF60 CRC32;

Query Match 88.2%; Score 45; DB 10; Length 330;
Best Local Similarity 62.5%; Pred. No. 9.41e-01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 234 SVPDPKAV 241
QY 264 SVPDPKSI 271

RESULT 3
ID Q42579 PRELIMINARY; PRT; 330 AA.
AC Q42579;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PEROXIDASE ATP1A PRECURSOR (EC 1.11.1.7) (MYELOPEROXIDASE).
GN YPR9.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
[1]
RN RP SEQUENCE FROM N.A.
RC TISSUE-STEM;
RX MEDLINE; 97238476.
RA KJAERGAARD I.V.H., JESPERSEN H.M., RASMUSSEN S.K., WELINDER K.G.;
RT "Sequence and RT-PCR expression analysis of two peroxidases from
RT Arabidopsis thaliana belonging to a novel evolutionary branch of plant
RT peroxidases."
RL Plant Mol. Biol. 33:699-708(1997).
CC -!- CATALYTIC ACTIVITY: DONOR + H(2)O(2) -> OXIDIZED DONOR + 2 H(2)O.
CC -!- COFACTOR: HEME.
DR EMBL; X98189; CAA66862.1; -.
DR HSSP; P22195; 1SCH.
DR MENDEL; 6559; Arath.Ypr9; 6559.
DR PFAM; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
KW Signal; Peroxidase; Oxidoreductase.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 330 POTENTIAL.
FT CHAIN 30 330 POTENTIAL.
SQ SEQUENCE 330 AA; 37383 MW; 250E7674 CRC32;

```

```

Query Match 88.2%; Score 45; DB 10; Length 330;
Best Local Similarity 62.5%; Pred. No. 9.41e-01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 234 SVPDPKAV 241
QY 264 SVPDPKSI 271

RESULT 4
ID P70083 PRELIMINARY; PRT; 991 AA.
AC P70083;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE SARCOPLASMIC/ENDOPLASMIC RETICULUM CALCIUM ATPASE 1A.
OS Makaira nigricans.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
OC Perciformes; Scombroidei; Xiphioidea; Istiophoridae; Makaira.
[1]
RN RP SEQUENCE FROM N.A.
RC TISSUE-FAST-TWITCH GLYCOLYTIC SKELETAL MUSCLE;
RA LONDRVILLE R.L., FRANK J.P.C.F., BLOCK B.A.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U65228; AAB08097.1; -.
DR PFAM; PF00122; E1-E2_ATPase; 1.
DR PRINTS; PR00119; CATATPASE.
DR SEQUENCE 991 AA; 108780 MW; 903F1E9F CRC32;

Query Match 86.3%; Score 44; DB 13; Length 991;
Best Local Similarity 62.5%; Pred. No. 1.78e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 193 SVPDPRAV 200
QY 264 SVPDPKSI 271

```

```

RESULT 5
ID P70084 PRELIMINARY; PRT; 996 AA.
AC P70084;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE SARCOPLASMIC/ENDOPLASMIC RETICULUM CALCIUM ATPASE 1B.
OS Makaira nigricans.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
OC Perciformes; Scombroidei; Xiphioidea; Istiophoridae; Makaira.
[1]
RN RP SEQUENCE FROM N.A.
RC TISSUE-EYE MUSCLE;
RA LONDRVILLE R.L., FRANK J.P.C.F., BLOCK B.A.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U65229; AAB08098.1; -.
DR PFAM; PF00122; E1-E2_ATPase; 1.
DR PRINTS; PR00119; CATATPASE.
DR SEQUENCE 996 AA; 109246 MW; 7519963E CRC32;

Query Match 86.3%; Score 44; DB 13; Length 996;
Best Local Similarity 62.5%; Pred. No. 1.78e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 193 SVPDPRAV 200
QY 264 SVPDPKSI 271

RESULT 6
ID Q9XU13 PRELIMINARY; PRT; 1004 AA.
AC Q9XU13;
DT 01-NOV-1999 (TREMBlrel. 12, Created)

```

[2]
RN SEQUENCE FROM N.A.
RP MEDLINE: 94150718.
RX
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAN S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SNALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

AC	Q9ZSH8;		
DT	01-MAY-1999	(TREMBlrel. 10,	Created)
DT	01-MAY-1999	(TREMBlrel. 10,	Last sequence update)
DT	01-MAY-1999	(TREMBlrel. 10,	Last annotation update)

DE T15B16.1 PROTEIN.
GN T15B16.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC eumyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA WASHU;
RT "The A. thaliana Genome Sequencing Project."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA STONEKING T., SMITH R.;
RT "The sequence of A. thaliana T15B16."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA WATERSTON R.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF104919; AAC72861.1; 88F096EE CRC32;
SQ SEQUENCE 220 AA; 25252 MW; 25252 MW; 88F096EE CRC32;
Query Match 84.3%; Score 43; DB 10; Length 220;
Best Local Similarity 75.0%; Pred. No. 3.35e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 30 TLPDPKSI 37
QY 264 SVPDPKSI 271
RESULT 10
ID O16657 PRELIMINARY; PRT; 287 AA.
AC O16657;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE C03H5.3 PROTEIN.
GN C03H5.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSKOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA JONES A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA ROHLFING T., WOHLDMANN P.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF016674; AAB66130.1;
SQ SEQUENCE 287 AA; 31998 MW; 4B4ED5ED CRC32;
Query Match 84.3%; Score 43; DB 5; Length 287;
Best Local Similarity 75.0%; Pred. No. 3.35e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 183 SVAEPKSI 190
QY 264 SVPDPKSI 271
RESULT 11
ID O64970 PRELIMINARY; PRT; 331 AA.
AC O64970;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-JAN-1999 (TREMBlrel. 09, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CATIONIC PEROXIDASE 2 (EC 1.11.1.7) (MYELOPEROXIDASE).
GN YPR9 OR PRX2.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eumyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, HARSOY 63; TISSUE=SEED COAT;
RA BOUTILLIER K., GIJZEN M., KUFU K., MIKI B.;
RT "Cationic peroxidase from immature soybean seeds."
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.
CC -1- COFACTOR: HEME.
DR EMBL: AF039027; AAC83463.1;
DR HSPF; P22195; ISCH.
DR MENDEL; 29532; Glyma; Ypr9; 29532.
DR PFAM; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
KW Peroxidase; Oxidoreductase.
SQ SEQUENCE 331 AA; 37204 MW; 298E460A CRC32;
Query Match 84.3%; Score 43; DB 10; Length 331;
Best Local Similarity 50.0%; Pred. No. 3.35e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 232 AIPDPKAV 239
QY 264 SVPDPKSI 271
RESULT 12
ID O9XFI6 PRELIMINARY; PRT; 336 AA.
AC O9XFI6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PEROXIDASE.
GN PRX2B.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eumyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, HARSOY 63; TISSUE=SEED COAT;
RA HARRIS N., GIJZEN M.;
RT "Soybean peroxidase transcripts."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF145348; AAD37374.1;
KW Peroxidase.
SQ SEQUENCE 336 AA; 37470 MW; F296B801 CRC32;

Query Match 84.3%; Score 43; DB 10; Length 336;
Best Local Similarity 50.0%; Pred. No. 3.35e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 236 AIPDPKAV 243
QY 264 SVPDPKSI 271

RESULT 13
ID 081799 PRELIMINARY; PRT; 503 AA.
AC 081799;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE HYPOTHETICAL 57.0 KD PROTEIN.
GN F8D20.190.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA KOETTER P., HEMPEL S., ENTIAN K.-D., HOEISEL J., JESSE T.,
RA HEIJNEN L., VOS P., MEWES H.W., MAYER K.F.X., SCHUELLER C., BEVAN M.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031133; CAA20038.1; -.
KW Hypothetical protein.
SQ SEQUENCE 503 AA; 56991 MW; FAED65A4 CRC32;

Query Match 84.3%; Score 43; DB 10; Length 503;
Best Local Similarity 75.0%; Pred. No. 3.35e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 355 TLPDPKSI 362
QY 264 SVPDPKSI 271

RESULT 14
ID 095256 PRELIMINARY; PRT; 599 AA.
AC 095256;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE INTERLEUKIN-18 RECEPTOR ACCESSORY PROTEIN-LIKE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99009050.
RA BORN T.L., THOMASSEN E., BIRD T.A., SIMS J.E.;
RT "Cloning of a novel receptor subunit, AcPL, required for interleukin-18 signaling."
RL J. Biol. Chem. 273:29445-29450(1998).
DR EMBL; AF077346; AAC72196.1; -.
KW Receptor.
SQ SEQUENCE 599 AA; 68309 MW; 8A1F3AEB CRC32;

Query Match 84.3%; Score 43; DB 4; Length 599;
Best Local Similarity 75.0%; Pred. No. 3.35e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

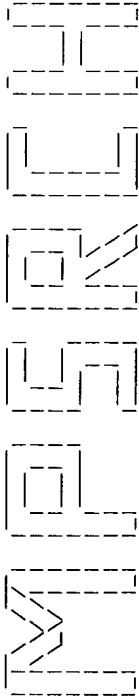
Db 301 SVPKSI 308
QY 264 SVPDPKSI 271

RESULT 15
ID 017315 PRELIMINARY; PRT; 1002 AA.
AC 017315;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE SARCO/ENDOPLASMIC RETICULUM CA2+-ATPASE.
GN SERCA.
OS Procambarus clarkii (Red swamp crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacoidea; Cambaridae; Procambarus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TAIL MUSCLE.
RA ZHANG Z., CHEN D., WHEATLY M.G.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF025849; AAB82291.1; -.
DR PFAM; PF00122; E1-E2_ATPase; 1.
DR PRINTS; PR00119; CATATPASE.
SQ SEQUENCE 1002 AA; 110315 MW; B49A57BC CRC32;

Query Match 84.3%; Score 43; DB 5; Length 1002;
Best Local Similarity 50.0%; Pred. No. 3.35e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 193 AIPDPKAV 200
QY 264 SVPDPKSI 271

Search completed: Wed May 10 13:48:10 2000
Job time : 244 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 13:57:16 2000; MasPar time 2.81 Seconds
Tabular output not generated. 67.400 Million cell updates/sec

Title: >US-09-376-430-2
Description: (286-293) from US09376430A.pep (21 of 25)
Perfect Score: 60
Sequence: 1 WITDTQNV 8

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
l:geneseqp

Statistics: Mean 15.923; Variance 57.458; scale 0.277

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	47	78.3	154	1 W77768	CDP-ribitol pyrophosph	3.39e+02
2	45	75.0	398	1 R37302	RGL precursor.	5.14e+02
3	45	75.0	1255	1 W2406	Human HER-2/neu oncogene	5.14e+02
4	45	75.0	1255	1 W0111	HER-2/neu protein.	5.14e+02
5	45	75.0	1433	1 R39568	Sequence of c-erbB-2 t	5.14e+02
6	44	73.3	415	1 W40536	Nucleotide sequence of	6.32e+02
7	44	73.3	619	1 R27651	Human calcium channel	6.32e+02
8	44	73.3	1967	1 R33547	Sequence of the alpha	6.32e+02
9	44	73.3	1968	1 R33139	Human calcium channel	6.32e+02
10	44	73.3	2138	1 R72607	Human neuronal calcium	6.32e+02
11	44	73.3	2161	1 W63149	Human calcium channel	6.32e+02
12	44	73.3	2161	1 R71002	Human neuronal calcium	6.32e+02
13	44	73.3	2161	1 R33545	Sequence of the alpha	6.32e+02
14	44	73.3	2161	1 R71001	Human neuronal calcium	6.32e+02
15	44	73.3	2161	1 W63137	Human calcium channel	6.32e+02
16	44	73.3	2163	1 R71003	Human neuronal calcium	6.32e+02
17	44	73.3	2366	1 W68388	Clostridium difficile	6.32e+02
18	44	73.3	2366	1 R50111	C. difficile toxin B.	6.32e+02
19	44	73.3	2516	1 W01875	Neuronal invertebrate	6.32e+02
20	44	73.3	2516	1 W01884	Invertebrate calcium c	6.32e+02
21	43	71.7	136	1 W88980	Polypeptide fragment e	7.77e+02
22	43	71.7	352	1 W68097	Chlamydomonas reinhard	7.77e+02
23	43	71.7	353	1 R68890	Japanese Black Pine ch	7.77e+02

24	43	71.7	364	1 W88979	Polypeptide fragment e	7.77e+02
25	43	71.7	501	1 W57331	Cytosolic glycerol-3-p	7.77e+02
26	43	71.7	501	1 W60259	Klebsiella pneumoniae	7.77e+02
27	43	71.7	584	1 W42385	Isolate OC9a alkaline	7.77e+02
28	43	71.7	1192	1 W57900	Protein of clone CO722	7.77e+02
29	43	71.7	1304	1 W59994	Human neural cell adhe	7.77e+02
30	42	70.0	15	1 R47683	HIV epitope #16.	9.53e+02
31	42	70.0	125	1 Y12858	Human 5' EST secreted	9.53e+02
32	42	70.0	166	1 R42017	Trans-sialidase/neuram	9.53e+02
33	42	70.0	166	1 R42014	Protein with trans-sia	9.53e+02
34	42	70.0	166	1 R42015	Trans-sialidase/neuram	9.53e+02
35	42	70.0	166	1 R42016	TCNA Trans-sialidase/n	9.53e+02
36	42	70.0	280	1 R44156	Moraxella bovis mboB g	9.53e+02
37	42	70.0	342	1 W23140	Meripilus giganteus ga	9.53e+02
38	42	70.0	468	1 W55675	H. pylori ORF hp6p124	9.53e+02
39	42	70.0	500	1 W55736	H. pylori ORF hp6p124	9.53e+02
40	42	70.0	555	1 R12339	Ascorbate oxidase from	9.53e+02
41	42	70.0	642	1 Y01541	Alpha(2-3) trans-siali	9.53e+02
42	42	70.0	1060	1 Y01540	Trypanosoma cruzi alph	9.53e+02
43	42	70.0	1256	1 R27746	Muramidase released pr	9.53e+02
44	42	70.0	3433	1 W22017	Utrophin.	9.53e+02
45	41	68.3	808	1 W94902	Mouse pheromone recept	1.17e+03

ALIGNMENTS

RESULT 1
ID W77768 standard; Protein; 154 AA.
AC W77768;
DE 30-OCT-1998 (first entry)
DT CDP-ribitol pyrophosphorylase protein.
KW Staphylococcus aureus protein; immune response induction; eye infection;
KW antibody production; T-cell immune response; gastrointestinal infection;
KW respiratory infection; inhibitor; bacterial infection; cardiac infection;
KW central nervous system; kidney infection; urinary tract infection;
KW antimicrobial compound identification; broad spectrum antibiotic;
KW therapy.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT Misc_difference 1..154 /note= "residues designated X are unspecified, and represented as xaa in the specification"
FT
PD EP-841394-A2.
PE 13-MAY-1998.
PF 24-SEP-1997; 307485.
PR 24-SEP-1996; US-027032.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Black MT, Burnham MKR, Hodgson JE, Knowles DJC,
PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,
PI Ward JM;
PI WPI: 98-252940/23.
DR N-PSDB; V53554.
DR New nucleic acid sequences from Staphylococcus aureus WCHU29 -
PT useful in vaccines and for treatment of bacterial infections of e.g.
PT respiratory tract and central nervous system
PS Claim 11; Page 383; 390pp; English.
CC This sequence represents a Staphylococcus aureus protein, that based on
CC homology with a Haemophilus influenzae protein, is a CDP-ribitol
CC pyrophosphorylase, and is encoded by a DNA sequence of the invention.
CC The DNA sequences were isolated from Staphylococcus aureus WCHU29
CC (NCIMB 40771). Host cells containing the DNA sequences are used to
CC produce polypeptides or fragments. The proteins are used in the treatment
CC of disease, for inducing an immune response by administering them, to
CC produce antibody and/or T-cell immune response. Antagonists of the
CC proteins are used for the inhibition of bacterial polypeptides.
CC Conditions which may be treated include bacterial infections, especially
CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,
CC urinary tract, skin, bones and joints. The proteins can also be used to
CC identify antimicrobial compounds which are broad spectrum antibiotics,
CC especially useful in the treatment of H. pylori infection.
SQ Sequence 154 AA;

```

Query Match      78.3%; Score 47; DB 1; Length 154;
Best Local Similarity 50.0%; Pred. No. 3.38e+02;
Matches      4; Conservative      3; Mismatches      1; Indels      0; Gaps      0;

Db      15 WISHTQDI 22
QY      286 WITDTQNV 293

RESULT      2
ID      R37302 standard; Protein; 398 AA.
AC      R37302;
DE      20-SEP-1993 (first entry)
KW      RGL precursor.
KW      Rabbit gastric lipase; RGL; PRGLN2.1; fat; bioconversion;
OS      hydrolysis; transesterification.
OS      Oryctolagus cuniculus.
FH      Key
FT      Location/Qualifiers
FT      23..398
FT      /note= "claim 1; page 10-11"
PN      EP-542629-A.
PD      19-MAY-1993.
PF      12-NOV-1992; 403055.
PR      13-NOV-1991; FR-013948.
PA      (LJOU ) INST RECH JOUVEINAL.
PI      Benicourt C, Blanchard C, Junien J;
DR      WPI; 93-161080/20.
DR      N-PSDB; Q42310.
PT      Rabbit gastric lipase, its precursor and their DNA - useful for
PT      treating conditions linked to gastric lipase deficiency, such as
PT      mucoviscidosis and pancreatic exocrine insufficiency
PS      Claim 1; Fig 7; 31pp; French.
CC      RGL, opt. used with other lipases, are useful therapeutically
CC      (1) to facilitate absorption of ingested fats in patients deficient
CC      in endogenous gastric lipase and (2) to treat disorders caused by
CC      inadequate lipase prodn. esp. mucoviscidosis or pancreatic exocrine
CC      deficiency. It can also be used (partic. when immobilised) for
CC      enzymatic bioconversion, e.g. hydrolysis or transesterification.
SQ      Sequence 398 AA;

Query Match      75.0%; Score 45; DB 1; Length 398;
Best Local Similarity 50.0%; Pred. No. 5.14e+02;
Matches      4; Conservative      3; Mismatches      1; Indels      0; Gaps      0;

Db      344 WLADPODV 351
QY      286 WITDTQNV 293

RESULT      3
ID      W92406 standard; Protein; 1255 AA.
AC      W92406;
DE      21-APR-1999 (first entry)
DE      Human HER-2/neu oncogene protein.
KW      HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW      malignancy; treatment; tumour.
OS      Homo sapiens.
FH      Key
FT      Location/Qualifiers
FT      Region
FT      676..1255
FT      /note= "region which elicits immune response".
PN      US5869445-A.
PD      09-FEB-1999.
PF      01-APR-1996; 625101.
PR      01-APR-1996; US-625101.
PR      17-MAR-1993; US-033644.
PR      12-AUG-1993; US-106112.
PR      31-MAR-1995; US-414417.
PA      (UNIW ) UNIV WASHINGTON.
PI      Cheever MA, Disis ML;
DR      WPI; 99-152835/13.
DR      N-PSDB; X01912.
PT      Use of HER-2/neu polypeptides - for eliciting an immune response to
PT      an HER-2/neu associated malignancy, particularly for treating or

```

```

PT      preventing tumours
PS      Claim 3; Column 31-38; 26pp; English.
CC      This sequence represents the human HER-2/neu oncogene protein. A fragment
CC      of this protein is used in a method for eliciting or enhancing an immune
CC      response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC      B cells to produce an immune response to the HER-2/neu protein. The
CC      method can be used for immunisation against a malignancy in which the
CC      HER-2/neu oncogene is associated and in the treatment of an existing
CC      tumour, or to prevent tumour occurrence or reoccurrence.
SQ      Sequence 1255 AA;

Query Match      75.0%; Score 45; DB 1; Length 1255;
Best Local Similarity 62.5%; Pred. No. 5.14e+02;
Matches      5; Conservative      1; Mismatches      2; Indels      0; Gaps      0;

Db      739 WIPDGENV 746
QY      286 WITDTQNV 293

RESULT      4
ID      W01111 standard; Protein; 1255 AA.
AC      W01111;
DE      01-JAN-1997 (first entry)
DE      HER-2/neu protein.
KW      HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW      breast cancer; Ovary cancer; colon cancer; lung cancer;
KW      prostate cancer; immunisation; tumour; vaccine; vector.
OS      Homo sapiens.
FH      Key
FT      Location/Qualifiers
FT      domain
FT      676..1255
FT      /label= Intracellular_domain
FT      /note= "claimed domain, useful for immunisation"
PN      W09630514-A1.
PD      03-OCT-1996.
PF      28-MAR-1996; U01689.
PR      31-MAR-1995; US-414417.
PA      (UNIW ) UNIV WASHINGTON.
PI      Cheever MA, Disis ML;
DR      WPI; 96-455361/45.
DR      N-PSDB; T40739.
PT      DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT      treatment of malignancies with which the HER-2/neu oncogene is
PT      associated
PS      Claim 2; Page 56-61; 71pp; English.
CC      Human HER-2/neu protein (W01111), also called p185 or c-erbB2, is
CC      the product of the HER-2/neu oncogene (see also T40739). The
CC      protein is over-expressed in various cancers, including breast,
CC      ovarian, colon, lung and prostate. The intracellular domain of the
CC      protein can be used to immunise an animal against a malignancy with
CC      which the oncogene is associated. The polypeptide can be produced
CC      in transformed host cells for use in immunisation. Alternatively,
CC      animal cells are transfected in vivo or ex vivo with a viral vector
CC      that directs expression of the polypeptide.
SQ      Sequence 1255 AA;

Query Match      75.0%; Score 45; DB 1; Length 1255;
Best Local Similarity 62.5%; Pred. No. 5.14e+02;
Matches      5; Conservative      1; Mismatches      2; Indels      0; Gaps      0;

Db      739 WIPDGENV 746
QY      286 WITDTQNV 293

RESULT      5
ID      R39568 standard; Protein; 1433 AA.
AC      R39568;
DE      07-FEB-1994 (first entry)
DE      Sequence of c-erbB-2 tumour antigen.
KW      Tumour antigen; c-erbB-2; glycoprotein.
OS      Homo sapiens.
PN      W09316185-A.

```


PI 19-AUG-1993.
PF 05-FEB-1993; U01055.
PR 06-FEB-1992; US-831967.
PA (CETU) CETUS ONCOLOGY CORP.
PA (CREA) CREATIVE BIOMOLECULES INC.
PI Houston LL, Huston JS, Oppermann H, Ring DB;
DR WPI; 93-272889/34.
DR N-PSDB; Q46083.
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
PS Disclosure; pages 48-54; 87pp; English.
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see Q46083, R39568). X in R39583 represents
CC the location of a stop codon in Q46083.
SQ Sequence 1433 AA;
Query Match 75.0%; Score 45; DB 1; Length 1433;
Best Local Similarity 62.5%; Pred. No. 5.14e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 739 WIPDGENV 746
QY 286 WITDTQNV 293
RESULT 6
ID W40536 standard; Protein; 415 AA.
AC W40536;
DT 09-SEP-1998 (first entry)
DE Nucleotide sequence of the xylitol oxidase protein.
KW Xylitol oxidase gene.
OS Streptomyces sp.
PN J10075785-A.
PD 24-MAR-1998.
PF 30-AUG-1996; 230952.
PR 30-AUG-1996; JP-230952.
PA (IKED-) IKEDA SHOKKEN KK.
PA (NIPK) NIPPON KAYAKU KK.
DR WPI; 98-244359/22.
DR N-PSDB; V11297.
PT DNA encoding xylitol oxidase - useful for production of enzyme on
PT large scale
PS Claim 4; Page 2; 18pp; Japanese.
CC This is the amino acid sequence of the xylitol oxidase protein, used
CC in the method of the invention, which involves the production of the
CC encoded (xylitol oxidase) enzyme on a large scale. Xylitol oxidase
CC gene was cloned to produce Escherichia coli JM109/pxY010, 020 and 030,
CC which was further used to prepare xylitol oxidase.
SQ Sequence 415 AA;
Query Match 73.3%; Score 44; DB 1; Length 415;
Best Local Similarity 50.0%; Pred. No. 6.32e+02;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 343 WVKDTEAV 350
QY 286 WITDTQNV 293
RESULT 7
ID R27651 standard; Protein; 619 AA.
AC R27651;
DT 03-MAR-1993 (first entry)
DE Human calcium channel 27980/13.
KW Plasmid pCA9.3; Ca-flux assay.
OS Homo sapiens.
PN EP-507170-A.
PD 07-OCT-1992.
PF 23-MAR-1992; 104970.
PR 04-APR-1991; DE-110785.
PA (FARB) BAYER AG.

PI Franz J, Rae P, Unterbeck A, Weingaertner B;
DR WPI; 92-333446/41.
DR N-PSDB; Q29271.
PT Cloned human neuronal calcium channel sub-types - useful in
PT calcium flux assays to screen for neurone-specific calcium
PT channel ligands
PS Claim 2; Page 83-85; 101pp; German.
CC Human neuroblastoma cell line, hippocampus, frontal and temporal
CC cortex and visual cortex cDNA banks were screened with a probe
CC containing carp skeletal muscle Ca-channel cDNA. The cDNA clone
CC pCA9.3 was sequenced. The 5' end of the clone begins directly after
CC Domain I (amino acid 337) and the 3' end is at amino acid 922. The
CC neuronal calcium channel protein can be used for screening for Ca
CC channel ligands (agonists or antagonists). See also Q29259-Q29275.
SQ Sequence 619 AA;
Query Match 73.3%; Score 44; DB 1; Length 619;
Best Local Similarity 37.5%; Pred. No. 6.32e+02;
Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
Db 32 WITQAEI 39
QY 286 WITDTQNV 293
RESULT 8
ID R33547 standard; Protein; 1967 AA.
AC R33547;
DT 30-JUN-1993 (first entry)
DE Sequence of the alpha 1C human calcium channel subunit.
KW Human calcium channel subunit; diagnosis; agonist; antagonist;
KW Lambert Eaton syndrome.
OS Homo sapiens.
PN W09304083-A.
PD 04-MAR-1993.
PF 14-AUG-1992; U06903.
PR 15-AUG-1991; US-745206.
PR 10-APR-1992; US-868354.
PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
PI Williams ME;
DR WPI; 93-093936/11.
DR N-PSDB; Q37813.
PT DNA encoding specific human calcium channel sub-units - used for
PT identifying calcium channel agonists and antagonists and
PT diagnosing Lambert Eaton syndrome
PS Disclosure; Page 102-109; 150pp; English.
CC Numerous alpha 1C-specific DNA clones were isolated.
CC Characterisation of the sequence revealed the alpha 1C coding
CC sequence, the alpha 1C initiation of translation sequence, and an
CC alternatively spliced region of alpha 1C. Q37814 and Q37815 encode
CC two possible amino terminal ends of the alpha 1C protein. Q37816
CC encodes an alternative exon for the IV S3 transmembrane domain.
SQ Sequence 1967 AA;
Query Match 73.3%; Score 44; DB 1; Length 1967;
Best Local Similarity 37.5%; Pred. No. 6.32e+02;
Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
Db 439 WITQAEI 446
QY 286 WITDTQNV 293
RESULT 9
ID W63139 standard; Protein; 1968 AA.
AC W63139;
DT 12-OCT-1998 (first entry)
DE Human calcium channel alpha-1C subunit.
KW Alpha-1C subunit; human; calcium channel; assay; detection;
KW characterisation; Lambert Eaton Syndrome; LES; diagnosis.
OS Homo sapiens.
FH Key Location/Qualifiers

FT Misc_difference 448 /note= "not specified"
 FT US5792846-A.
 PN 11-AUG-1998.
 PD 31-MAY-1995; 455543.
 PF 04-APR-1994; US-223305.
 PR 04-APR-1988; US-176899.
 PR 04-APR-1989; US-603751.
 PR 04-APR-1989; WO-001408.
 PR 20-FEB-1990; US-482384.
 PR 30-NOV-1990; US-620250.
 PR 15-AUG-1991; US-745206.
 PR 31-MAY-1995; US-455543.
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
 PI Williams ME;
 DR WPI: 98-456192/39.
 DR N-PSDB: V42681.
 DR DNA encoding human calcium channel alpha 1B subunit protein -
 PT useful for recombinant production of the channel for screening of
 PT its modulators, and diagnosis of Lambert Eaton Syndrome
 PS Example 1: Columns 229-238; 166pp; English.
 CC The present sequence represents the alpha-1C subunit of a human calcium
 CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
 CC that allow controlled entry of calcium ions into cells. This leads
 CC to depolarisation events required for muscle contraction. The recombinant
 CC subunit, when expressed with nucleic acids encoding the complete calcium
 CC channel, can be used in assays for the detection and characterisation of
 CC compounds that modulate the channel. The DNA encoding the subunits can
 CC be alternatively spliced when transcribed, giving more than one form of
 CC the protein from the same transcript, each having slightly different
 CC properties. In addition, the reactivity of the alpha 1 subunit with IgG
 CC molecules from the serum of an individual with Lambert Eaton Syndrome
 CC (LES) can be used as a diagnostic for the disease.
 SQ Sequence 1968 AA;
 Query Match 73.3%; Score 44; DB 1; Length 1968;
 Best Local Similarity 37.5%; Pred. No. 6.32e+02;
 Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
 Db 440 WITQAEADI 447
 Qy 286 WITDTQNV 293
 RESULT 10
 ID R72607 standard; Protein; 2138 AA.
 AC R72607;
 DT 01-DEC-1995 (first entry)
 DE Human neuronal calcium channel subunit alpha 1c-2.
 KW Calcium channel subunit; antagonist; agonist; diagnosis;
 KW Lambert Eaton Syndrome.
 OS Homo sapiens.
 PN W09504822-A.
 PD 16-FEB-1995.
 PF 11-AUG-1994; U09230.
 PR 11-AUG-1993; US-105536.
 PR 05-NOV-1993; US-149097.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;
 DR WPI: 98-090900/12.
 DR N-PSDB: Q87834.
 DR DNA encoding human calcium channel sub-unit(s) - used for
 PT developing prods. for studying calcium channels, e.g. for
 PT obtaining agonists and antagonists
 PS Claim 34; Page 260-269; 285pp; English.
 CC Numerous alpha 1c-specific cDNA clones were isolated in order to
 CC characterise the alpha 1c coding sequence, the initiation of
 CC translation and an alternatively spliced region. Q84655 sets
 CC forth one alpha 1c coding sequence (alpha 1c-1) and R71003 sets
 CC out its deduced AA sequence. Q87834 and R72607 set out another
 CC splice variant, designated alpha 1c-2. Q84656 encodes an
 CC alternative exon for the IV S3 transmembrane domain. Other

CC alpha 1c variants can be constructed by selecting alternative
 CC amino terminal ends in place of the ends in Q84655 and Q87834
 CC and/or inserting the alternative exon in the appropriate
 CC location (see Q84655 FT). In addition, a nt. sequence (see Q84655
 CC FT) can be deleted or inserted to produce an alternative alpha 1c
 CC splice variant.
 SQ Sequence 2138 AA;
 Query Match 73.3%; Score 44; DB 1; Length 2138;
 Best Local Similarity 37.5%; Pred. No. 6.32e+02;
 Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
 Db 440 WITQAEADI 447
 Qy 286 WITDTQNV 293
 RESULT 11
 ID W63149 standard; Protein; 2161 AA.
 AC W63149;
 DT 12-OCT-1998 (first entry)
 DE Human calcium channel alpha-1D subunit.
 KW Alpha-1D subunit; human; calcium channel; assay; detection;
 KW Characterisation; Lambert Eaton Syndrome; LES; diagnosis.
 OS Homo sapiens.
 PN US5792846-A.
 PD 11-AUG-1998.
 PF 31-MAY-1995; 455543.
 PR 04-APR-1994; US-223305.
 PR 04-APR-1988; US-176899.
 PR 04-APR-1989; US-603751.
 PR 04-APR-1989; WO-001408.
 PR 20-FEB-1990; US-482384.
 PR 30-NOV-1990; US-620250.
 PR 15-AUG-1991; US-745206.
 PR 31-MAY-1995; US-455543.
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
 PI Williams ME;
 DR WPI: 98-456192/39.
 DR N-PSDB: V42697.
 DR DNA encoding human calcium channel alpha 1B subunit protein -
 PT useful for recombinant production of the channel for screening of
 PT its modulators, and diagnosis of Lambert Eaton Syndrome
 PS Disclosure; Columns 271-284; 166pp; English.
 CC The present sequence represents the alpha-1D subunit of a human calcium
 CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
 CC that allow controlled entry of calcium ions into cells. This leads
 CC to depolarisation events required for muscle contraction. The recombinant
 CC subunit, when expressed with nucleic acids encoding the complete calcium
 CC channel, can be used in assays for the detection and characterisation of
 CC compounds that modulate the channel. The DNA encoding the subunits can
 CC be alternatively spliced when transcribed, giving more than one form of
 CC the protein from the same transcript, each having slightly different
 CC properties. In addition, the reactivity of the alpha 1 subunit with IgG
 CC molecules from the serum of an individual with Lambert Eaton Syndrome
 CC (LES) can be used as a diagnostic for the disease.
 SQ Sequence 2161 AA;
 Query Match 73.3%; Score 44; DB 1; Length 2161;
 Best Local Similarity 37.5%; Pred. No. 6.32e+02;
 Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
 Db 441 WITQAEADI 448
 Qy 286 WITDTQNV 293
 RESULT 12
 ID R71002 standard; Protein; 2161 AA.
 AC R71002;
 DT 30-NOV-1995 (first entry)
 DE Human neuronal calcium channel subunit alpha 1D including alternative.

DE exon encoding the IS6 transmembrane domain.
 KW Calcium channel subunit; antagonist; agonist; diagnosis;
 KW Lambert Eaton Syndrome.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_difference 373...406
 FT /label= encoded by alternative exon
 PN WO9504822-A.
 PD 16-FEB-1995.
 PF 11-AUG-1994; U09230.
 PR 11-AUG-1993; US-105536.
 PR 03-NOV-1993; US-149097.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Ellis SB, Gillespie A, Harpold MM, Mccue AF, Williams ME;
 DR WPI: 95-090900/12.
 DR N-PSDB: Q84654.
 PT DNA encoding human calcium channel sub-unit(s) - used for
 PT developing prods. for studying calcium channels, e.g. for
 PT obtaining agonists and antagonists
 PS Disclosure; Page 126-127; 285pp; English.
 CC The alpha 1D subunit cDNA has been isolated using fragments of the
 CC rabbit skeletal muscle calcium channel alpha 1 subunit cDNA as a
 CC probe to screen a cDNA library of human neuroblastoma cell line
 CC IMR32, to obtain clone alpha1.36. This clone was used as a probe to
 CC screen additional IMR32 cell cDNA libraries to obtain overlapping
 CC clones, which were then employed for screening until a sufficient
 CC series of clones to span the length of the nt sequence encoding the
 CC human alpha 1D subunit was obt'd. Full-length clones were then
 CC constructed by ligating partial clones. Q84653 shows the nt sequence
 CC of the cDNA encoding the alpha 1D subunit. The Alpha 1D protein has
 CC a calculated Mr of 245,163. It contains four putative internal
 CC repeated sequence regions which represent 24 putative transmembrane
 CC segments. It mediates DHP-sensitive high-voltage, long-lasting
 CC calcium channel activity. Q84654 shows an alternative exon encoding
 CC the IS6 transmembrane domain. The difference occurs in AAs 373-
 CC 406.
 SQ Sequence 2161 AA;
 Query Match 73.3%; Score 44; DB 1; Length 2161;
 Best Local Similarity 37.5%; Pred. No. 6.32e+02;
 Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
 Db 441 WITQAEI 448
 QY 286 WITDTQNV 293
 RESULT 13
 ID R33545 standard; Protein; 2161 AA.
 AC R33545;
 DT 30-JUN-1993 (first entry)
 DE Sequence of the alpha 1D human calcium channel subunit.
 KW Human calcium channel subunit; diagnosis; agonist; antagonist;
 KW Lambert Eaton Syndrome.
 OS Homo sapiens.
 PN WO9304083-A.
 PD 04-MAR-1993.
 PF 14-AUG-1992; U06903.
 PR 15-AUG-1991; US-745206.
 PR 10-APR-1992; US-868354.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, Mccue AF,
 PI Williams ME;
 DR WPI: 93-093936/11.
 DR N-PSDB: Q37811.
 PT DNA encoding specific human calcium channel sub-units - used for
 PT identifying calcium channel agonists and antagonists and
 PT diagnosing Lambert Eaton syndrome
 PS Disclosure; Page 93-101; 150pp; English.
 CC The alpha 1D subunit cDNA was isolated using fragments of the
 CC rabbit skeletal muscle calcium channel alpha 1 subunit cDNA as a
 CC probe to screen a cDNA library of a human neuroblastoma cell line,
 CC IMR32, to obtain clone alpha 1.36. This clone was used as a probe to

CC screen additional IMR32 cell cDNA libraries to obtain overlapping
 CC clones which were then employed for screening until a sufficient
 CC series of clones to span the length of the nucleotide sequence
 CC encoding the human alpha 1D subunit were obtained, see Q37811.
 CC Q37812 provides the sequence of an alternative exon encoding the IS6
 CC transmembrane domain of the alpha 1D subunit. The alpha 1D protein
 CC has a calculated Mr of 245,163.
 SQ Sequence 2161 AA;
 Query Match 73.3%; Score 44; DB 1; Length 2161;
 Best Local Similarity 37.5%; Pred. No. 6.32e+02;
 Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
 Db 441 WITQAEI 448
 QY 286 WITDTQNV 293
 RESULT 14
 ID R71001 standard; Protein; 2161 AA.
 AC R71001;
 DT 30-NOV-1995 (first entry)
 DE Human neuronal calcium channel subunit alpha 1D.
 KW Calcium channel subunit; antagonist; agonist; diagnosis;
 KW Lambert Eaton Syndrome.
 OS Homo sapiens.
 PN WO9504822-A.
 PD 16-FEB-1995.
 PF 11-AUG-1994; U09230.
 PR 11-AUG-1993; US-105536.
 PR 05-NOV-1993; US-149097.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Ellis SB, Gillespie A, Harpold MM, Mccue AF, Williams ME;
 DR WPI: 95-090900/12.
 DR N-PSDB: Q84653.
 PT DNA encoding human calcium channel sub-unit(s) - used for
 PT developing prods. for studying calcium channels, e.g. for
 PT obtaining agonists and antagonists
 PS Disclosure; Page 116-126; 285pp; English.
 CC The alpha 1D subunit cDNA has been isolated using fragments of the
 CC rabbit skeletal muscle calcium channel alpha 1 subunit cDNA as a
 CC probe to screen a cDNA library of human neuroblastoma cell line
 CC IMR32, to obtain clone alpha1.36. This clone was used as a probe to
 CC screen additional IMR32 cell cDNA libraries to obtain overlapping
 CC clones, which were then employed for screening until a sufficient
 CC series of clones to span the length of the nt sequence encoding the
 CC human alpha 1D subunit was obt'd. Full-length clones were then
 CC constructed by ligating partial clones. Q84653 shows the nt sequence
 CC of the cDNA encoding the alpha 1D subunit. The Alpha 1D protein has
 CC a calculated Mr of 245,163. It contains four putative internal
 CC repeated sequence regions which represent 24 putative transmembrane
 CC segments. It mediates DHP-sensitive high-voltage, long-lasting
 CC calcium channel activity.
 SQ Sequence 2161 AA;
 Query Match 73.3%; Score 44; DB 1; Length 2161;
 Best Local Similarity 37.5%; Pred. No. 6.32e+02;
 Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
 Db 441 WITQAEI 448
 QY 286 WITDTQNV 293
 RESULT 15
 ID W63137 standard; Protein; 2161 AA.
 AC W63137;
 DT 12-OCT-1998 (first entry)
 DE Human calcium channel alpha-1D subunit.
 KW Alpha-1D subunit; human; calcium channel; assay; detection;
 KW characterization; Lambert Eaton Syndrome; LES; diagnosis.
 OS Homo sapiens.
 PN US5792846-A.

PD 11-AUG-1998. 455543.
 PF 31-MAY-1995; US-223305.
 PR 04-APR-1994; US-176899.
 PR 04-APR-1988; US-603751.
 PR 04-APR-1989; WO-U01408.
 PR 04-APR-1989; WO-U01408.
 PR 20-FEB-1990; US-482384.
 PR 30-NOV-1990; US-620250.
 PR 15-AUG-1991; US-745206.
 PR 31-MAY-1995; US-455543.
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
 PI Williams NE;
 DR WPI; 98-456192/39.
 DR N-PSDB; V42679.
 PT DNA encoding human calcium channel alpha 1B subunit protein -
 PT useful for recombinant production of the channel for screening of
 PT its modulators, and diagnosis of Lambert Eaton Syndrome
 PS Example 1; Columns 261-272; 166pp; English.
 CC The present sequence represents the alpha-1B subunit of a human calcium
 CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
 CC that allow controlled entry of calcium ions into cells. This leads
 CC to depolarisation events required for muscle contraction. The recombinant
 CC subunit, when expressed with nucleic acids encoding the complete calcium
 CC channel, can be used in assays for the detection and characterisation of
 CC compounds that modulate the channel. The DNA encoding the subunits can
 CC be alternatively spliced when transcribed, giving more than one form of
 CC the protein from the same transcript, each having slightly different
 CC properties. In addition, the reactivity of the alpha 1 subunit with IgG
 CC molecules from the serum of an individual with Lambert Eaton Syndrome
 CC (LES) can be used as a diagnostic for the disease.
 SQ Sequence 2161 AA;

Query Match 73.3%; Score 44; DB 1; Length 2161;
 Best Local Similarity 37.5%; Pred. NO. 6.32e+02;
 Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 441 WITQAEI 448

QY 286 WITDQNV 293

Search completed: Wed May 10 13:57:24 2000
 Job time : 8 secs.

W P S R E H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 13:15:00 2000; Maspar time 3.91 Seconds
Tabular output not generated. 96.497 Million cell updates/sec

Title: >US-09-376-430-2
Description: (139-146) from US09376430A.pep (16 of 25)
Perfect Score: 55
Sequence: 1 SDLSYGD L 8

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 22.739; Variance 26.965; scale 0.843

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	49	89.1	281	2	F69277 branched-chain amino	6.19e+00
2	49	89.1	417	2	S19373 hypothetical protein	6.19e+00
3	48	87.3	284	2	E71174 hypothetical protein	1.00e+01
4	48	87.3	504	2	S00390 Ig gamma chain (clone	1.00e+01
5	48	87.3	506	2	A54190 cerebroside-sulfatase	1.00e+01
6	48	87.3	507	1	KJHUAH probable ABC-type tra	1.00e+01
7	48	87.3	530	2	D64819 hypothetical protein	1.00e+01
8	48	87.3	389	2	S69711 hypothetical protein	1.00e+01
9	47	85.5	377	2	F69172 adenosylmethionine de	1.61e+01
10	47	85.5	400	2	T03947 alpha-aminoacyl-pyl-cys	1.61e+01
11	47	85.5	3746	1	YGPLV3 alpha-aminoacyl-pyl-cys	1.61e+01
12	47	85.5	3791	1	YGPLV8 probable adenosylmeth	2.58e+01
13	46	83.6	319	2	S73159 coat protein - arabis	2.58e+01
14	46	83.6	398	2	T04059 RNA-2 polypeptide - a	2.58e+01
15	46	83.6	506	2	S13770 probable Acetyl-CoA s	4.09e+01
16	46	83.6	737	2	FQ0219 conserved hypothetical	4.09e+01
17	45	81.8	78	2	F72495 conserved hypothetical	4.09e+01
18	45	81.8	249	2	B69693 conserved hypothetical	4.09e+01
19	45	81.8	319	2	S78332 conserved hypothetical	4.09e+01
20	45	81.8	320	2	S75838 conserved hypothetical	4.09e+01
21	45	81.8	406	1	JC4996 biphenyl dioxygenase	4.09e+01
22	45	81.8	537	2	S25181 heat shock protein gr	4.09e+01
23	45	81.8	539	2	F70737 chaperonin groEL1 - M	4.09e+01

24	45	81.8	757	2	I45956 polymeric immunoglobu	4.09e+01
25	45	81.8	757	1	S48411 secretory component p	4.09e+01
26	44	80.0	149	2	S73865 MG211 homolog H10_orf	6.43e+01
27	44	80.0	302	2	H64026 hypothetical protein	6.43e+01
28	44	80.0	331	2	S77198 hypothetical protein	6.43e+01
29	44	80.0	340	2	E72739 conserved hypothetical	6.43e+01
30	44	80.0	366	2	G71067 conserved hypothetical	6.43e+01
31	44	80.0	386	2	A75077 conserved hypothetical	6.43e+01
32	44	80.0	386	2	A75077 iron (iii) abc transp	6.43e+01
33	44	80.0	532	2	H69303 iron (iii) abc transp	6.43e+01
34	44	80.0	532	2	S77656 long chain fatty acid	6.43e+01
35	44	80.0	539	2	A58538 4-hydroxybenzoate--Co	6.43e+01
36	44	80.0	647	2	S75116 squalene-hopene-cycla	6.43e+01
37	44	80.0	869	2	A71400 probable disease resi	6.43e+01
38	44	80.0	934	2	T05201 hypothetical protein	6.43e+01
39	44	80.0	956	1	Q0BEK2 UL105 protein - human	6.43e+01
40	44	80.0	1242	2	JS0670 insulin receptor subs	6.43e+01
41	44	80.0	2026	1	OYBY adenylate cyclase (EC	6.43e+01
42	44	80.0	2448	2	S53999 pyoverdine synthetase	6.43e+01
43	44	80.0	4466	1	S17231 dynein beta heavy cha	6.43e+01
44	43	78.2	232	2	E75157 aspartate racemase PA	1.00e+02
45	43	78.2	571	2	E64503 arylsulfatase homolog	1.00e+02
			879	2	B70014 antibiotic synthetase	1.00e+02

ALIGNMENTS

RESULT 1

ENTRY F69277 #type complete
TITLE branched-chain amino acid ABC transporter, ATP-binding
protein (braF-1) homolog - Archaeoglobus fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 28-May-1999

ACCESSIONS F69277

REFERENCE A69250

#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, P.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

#journal Nature (1997) 390:364-370

#title The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.

#cross-references PMID:98049343

#accession F69277

#status preliminary; nucleic acid sequence not shown;

#molecule_type DNA translation not shown

#residues 1-281 #label KLE

#cross-references GB:AE001090; GB:AE000782; NID:92689413; PID:92650419; TIGR:AF0222

CLASSIFICATION #superfamily ATP-binding cassette homology

KEYWORDS ATP; P-loop

FEATURE 71-252 #domain ATP-binding cassette homology #label ABC\

88-95 #region nucleotide-binding motif A (P-loop)

SUMMARY #length 281 #molecular-weight 31093 #checksum 903

Query Match. 89.1%; Score 49; DB 2; Length 281;

Best Local Similarity 62.5%; Pred. No. 6.19e+00;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 175 ANLSYSEL 182

Qy 139 SDLSYDGL 146

```

2
RESULT 2
ENTRY 2
TITLE 2
ALTERNATE_NAMES 2
ORGANISM 2
DATE 2
ACCESSIONS 2
REFERENCE 2
#authors 2
#submission 2
#accession 2
#molecule_type 2
#residues 2
#cross-references 2
REFERENCE 2
#authors 2
#journal 2
#title 2
#cross-references 2
#accession 2
#molecule_type 2
#residues 2
#cross-references 2
GENETICS 2
#map_position 2
SUMMARY 2
Query Match 2
Best Local Similarity 2
Matches 2
Db 2
Qy 2
RESULT 3
ENTRY 3
TITLE 3
ALTERNATE_NAMES 3
ORGANISM 3
DATE 3
ACCESSIONS 3
REFERENCE 3
#authors 3
#journal 3
#title 3
#cross-references 3
#accession 3
#molecule_type 3
#residues 3
#cross-references 3
#status 3
#molecule_type 3
#residues 3
#cross-references 3
#authors 3
#submission 3
#accession 3

```

this accession replaces an interim accession for a sequence replaced by GenBank

PH0593
#length 284 #molecular-weight 31563 #checksum 1941

Query Match 87.3%; Score 48; DB 2; Length 284;
Best Local Similarity 85.7%; Pred. No. 1.00e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 141 NLSYDGL 147
Qy 140 DLSYDGL 146

RESULT 4
ENTRY 4
TITLE 4
ALTERNATE_NAMES 4
ORGANISM 4
DATE 4
ACCESSIONS 4
REFERENCE 4
#authors 4
#journal 4
#title 4
#cross-references 4
#accession 4
#molecule_type 4
#residues 4
#cross-references 4
#note 4

S00390 #type fragment
Ig gamma chain (clone 36) - chicken (fragment)
Ig nu chain
#formal_name Gallus gallus #common_name chicken
07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 16-Aug-1996
S00390
Parvari, R.; Avivi, A.; Lentner, F.; Ziv, E.; Tel-Or, S.; Burstein, Y.; Schechter, I.
EMBO J. (1988) 7:739-744
Chicken immunoglobulin gamma-heavy chains: limited VH gene repertoire, combinatorial diversification by D gene segments and evolution of the heavy chain locus.
#cross-references MUID:88283642
#accession S00390
#molecule_type mRNA
#residues 1-504 #label PAR
#cross-references EMBL:X07174
#note this sequence was determined from the differentiated gene

KEYWORDS
SUMMARY
immunoglobulin
#length 504 #checksum 7605

Query Match 87.3%; Score 48; DB 2; Length 504;
Best Local Similarity 85.7%; Pred. No. 1.00e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 54 DLGYDGL 60
Qy 140 DLSYDGL 146

RESULT 5
ENTRY 5
TITLE 5
ALTERNATE_NAMES 5
ORGANISM 5
DATE 5
ACCESSIONS 5
REFERENCE 5
#authors 5
#journal 5
#title 5
#cross-references 5
#accession 5
#status 5
#molecule_type 5
#residues 5
#cross-references 5
#authors 5
#submission 5
#accession 5

A54190 #type complete
cerebroside-sulfatase (EC 3.1.6.8) precursor - mouse
arylsulfatase A
#formal_name Mus musculus #common_name house mouse
18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999
A54190; S34293; A38075; S34294
Kreysing, J.; Polten, A.; Hess, B.; von Figura, K.; Menz, K.; Steiner, F.; Gieselmann, V.
Genomics (1994) 19:249-256
Structure of the mouse arylsulfatase A gene and cDNA.
#cross-references MUID:94245194
#accession A54190
#status preliminary
#molecule_type mRNA
#residues 1-506 #label KRE
#cross-references GB:X73230; NID:g312271; PIDN:CAAS1702.1; PID:g312272
S34293
Kreysing, J.; Polten, A.; Hess, B.; Menz, K.; Steiner, F.; Gieselmann, V.
submitted to the EMBL Data Library, June 1993
#accession S34293

```

##status preliminary
##molecule_type DNA
##residues 1-506 ##label KR2
##cross-references EMBL:X73231; NID:g312268; PIDN:CAA51703.1;
PID:g312269

REFERENCE
A38075
#authors Grompe, M.; Pieretti, M.; Caskey, C.T.; Ballabio, A.
#journal Genomics (1992) 12:755-760
#title The sulfatase gene family: cross-species PCR cloning using
the MOPAC technique.
#cross-references MUID:92241876
#accession A38075

##molecule_type mRNA
##residues 27-72 ##label GRO
##cross-references GB:M82876; NID:g192027; PIDN:AAA37260.1; PID:g192028
##note sequence extracted from NCBI backbone (NCBIN:98744,
NCBIP:98749)
##note the protein sequence from Fig. 3 includes conserved
regions used in the primer
CLASSIFICATION #superfamily animal sulfatase
KEYWORDS glycoprotein; sulfuric ester hydrolase
FEATURE
1-17 #domain signal sequence #status predicted #label SIG\
18-506 #product cerebroside-sulfatase #status predicted #label
MAR\
68 #modified_site 3-oxoalanine (Cys) #status predicted\
157,183,349 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 506 #molecular-weight 53776 #checksum 7525

Query Match 87.3%; Score 48; DB 2; Length 506;
Best Local Similarity 85.7%; Pred. No. 1.00e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 29 DLGYGDL 35
QY 140 DLSYGDL 146

RESULT 6
ENTRY KJHUA #type complete
TITLE cerebroside-sulfatase (EC 3.1.6.8) precursor - human
ALTERNATE_NAMES arylsulfatase A (ASA); arylsulfatase (EC 3.1.6.1)
(misidentification)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Nov-1993 #sequence_revision 27-Oct-1995 #text_change
11-Jun-1999
ACCESSIONS S11031; G02857; A32207; S23932
REFERENCE
#authors Kreysing, J.; von Figura, K.; Gieselmann, V.
#journal Eur. J. Biochem. (1990) 191:627-631
#title Structure of the arylsulfatase A gene.
#cross-references MUID:90361046
#accession S11031
##molecule_type DNA
##residues 1-507 ##label KRE
##cross-references EMBL:X52150; NID:g28859; PIDN:CAA36398.1; PID:g28860
REFERENCE H01749
#authors Adams, M.D.; Kerlavage, A.R.; Fuldner, R.A.; Phillips, C.A.;
Venter, J.C.
#submission submitted to the EMBL Data Library, June 1996
#accession G02857
##status translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 'MS', 1-507 ##label ADA
##cross-references EMBL:U62317; NID:g1399959; PID:g13
##note an incorrect initiation codon was used
REFERENCE A32207
#authors Stein, C.; Gieselmann, V.; Kreysing, J.; Schmidt, B.;
Pohlmann, R.; Waheed, A.; Meyer, H.E.; O'Brien, J.S.; von
Figura, K.
#journal J. Biol. Chem. (1989) 264:1252-1259
#title Cloning and expression of human arylsulfatase A.

#cross-references MUID:89093115
#accession A32207
##molecule_type mRNA
##residues 1-358,'RPPAAGHROPBSAVSLLLPVLRPPGPGWGCADKWKVQSLHP',
402-507 ##label STE
##cross-references GB:X52151; GB:J04442; GB:J04593
##note parts of this sequence, including the amino end of the
mature protein, were determined by protein sequencing
REFERENCE S23932
#authors Fujii, T.; Kobayashi, T.; Honke, K.; Gasa, S.; Ishikawa, M.;
Shimizu, T.; Makita, A.
#journal Biochim. Biophys. Acta (1992) 1122:93-98
#title Proteolytic processing of human lysosomal arylsulfatase A.
#cross-references MUID:9338230
#accession S23932
##molecule_type protein
##residues 20-29,31,33;434-479 ##label FUJ
##experimental_source placenta
REFERENCE A57113
#authors Schmidt, B.; Selmer, T.; Ingendoh, A.; von Figura, K.
#journal Cell (1995) 82:271-278
#title A novel amino acid modification in sulfatases that is
defective in multiple sulfatase deficiency.
#cross-references MUID:9354208
#contents annotation; identification of 3-oxoalanine,
2-amino-3-oxopropanoic acid
GENETICS
#gene GDB:ARSA
##cross-references GDB:119007; OMIM:250100
#map_position 22q13.31-22qter
#introns 73/2; 153/3; 226/3; 283/2; 325/1; 367/3; 402/1
#note defects in this gene can cause metachromatic leukodystrophy
FUNCTION
#description hydrolyzes cerebroside 3-sulfate to release sulfate; can also
hydrolyze galactose-3-sulfate, ascorbate 2-sulfate, and
many phenol sulfates
CLASSIFICATION #superfamily animal sulfatase
KEYWORDS glycoprotein; lysosomal storage disease; lysosome; sulfuric
ester hydrolase
FEATURE
1-19 #domain signal sequence #status predicted #label SIG\
20-507 #product cerebroside-sulfatase #status predicted #label
MAR\
20-444,445-507 #product cerebroside-sulfatase component b #status
predicted #label MCB\
20-444,448-507 #product cerebroside-sulfatase component c #status
predicted #label MCC\
20-444,449-507 #product cerebroside-sulfatase minor component c #status
predicted #label MCD\
69 #modified_site 3-oxoalanine (Cys) #status experimental\
158 #binding_site carbohydrate (Asn) (covalent) #status
experimental\
184,350 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 507 #molecular-weight 53588 #checksum 6235

Query Match 87.3%; Score 48; DB 1; Length 507;
Best Local Similarity 85.7%; Pred. No. 1.00e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 30 DLGYGDL 36
QY 140 DLSYGDL 146

RESULT 7
ENTRY D64819 #type complete
TITLE probable ABC-type transport protein ybit - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
24-Oct-1998
ACCESSIONS D64819
REFERENCE A64720

```

```

#authors      Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
              Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
              Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
              Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
              Y.
#journal      Science (1997) 277:1453-1462
#title       The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession   D64819
#status      nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues    1-530 #label BLAT
#cross-references GB:AE000184; GB:U00096; NID:gl787036; PID:gl787041;
              UMGPIB0820
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene        Ybit
CLASSIFICATION #superfamily ATP-binding cassette homology
KEYWORDS      ATP; P-loop
FEATURE
17-228        #domain ATP-binding cassette homology #label ABC1\
34-41         #region nucleotide-binding motif A (P-loop)\
335-511       #domain ATP-binding cassette homology #label ABC2\
353-359       #region nucleotide-binding motif A (P-loop)
SUMMARY       #length 530 #molecular-weight 59858 #checksum 9156
Query Match   87.3%; Score 48; DB 2; Length 530;
Best Local Similarity 62.5%; Pred. No. 1.00e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 223 ADLDYDGL 230
QY 139 SDSLGYDGL 146

RESULT 8
ENTRY  #S69711 #type complete
TITLE  hypothetical protein YDR430c - yeast (Saccharomyces
ORGANISM cerevisiae)
DATE    22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change
ACCESSIONS S69711; S31321
REFERENCE   Dietrich, F.S.
#authors    submitted to the EMBL Data Library, August 1995
#description The sequence of S. cerevisiae lambda 3641 and cosmids 9461,
              9831, and 9410.
#accession  S69711
#molecule_type DNA
#residues   1-989 #label DIE
#cross-references EMBL:033007; NID:g927685; PID:g927711; MIPS:YDR430c
REFERENCE   Ellis, E.M.; Reid, G.A.
#authors    submitted to the EMBL Data Library, February 1993
#description The MTS1 gene from Saccharomyces cerevisiae encodes a
              putative RNA-binding protein involved in mitochondrial
              protein targeting.
#accession  S31321
#molecule_type DNA
#residues   'MPKHPRLA', 19-166 #label ELL
#cross-references EMBL:X70951; NID:g288585; PID:g288586
GENETICS
#map_position 4R
#note         YDR430c
SUMMARY       #length 989 #molecular-weight 112179 #checksum 2840
Query Match   87.3%; Score 48; DB 2; Length 989;
Best Local Similarity 75.0%; Pred. No. 1.00e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 220 TDLRYDGL 227
              :||| ||||

```

```

QY 139 SDSLGYDGL 146

RESULT 9
ENTRY  #F69172 #type complete
TITLE  hypothetical protein MTH55 - Methanobacterium
ORGANISM thermoautotrophicum (strain Delta H)
DATE    05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
              05-Jun-1998
ACCESSIONS F69172
REFERENCE   A69000
#authors    Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
              Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
              Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
              Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, P.;
              Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; Caruso,
              A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
              McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.;
              Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
              J.; Reeve, J.N.
#journal    J. Bacteriol. (1997) 179:7135-7155
#title      Complete genome sequence of Methanobacterium
              thermoautotrophicum Delta H: functional analysis and
              comparative genomics.
#cross-references MUID:98037514
#accession  F69172
#status     preliminary; nucleic acid sequence not shown;
              translation not shown
#molecule_type DNA
#residues   1-377 #label MTH
#cross-references GB:AE000797; GB:AE000666; NID:g2621083; PID:g2621091
#experimental_source strain Delta H
GENETICS
#gene       MTH55
SUMMARY     #length 377 #molecular-weight 42155 #checksum 1764
Query Match 85.5%; Score 47; DB 2; Length 377;
Best Local Similarity 62.5%; Pred. No. 1.61e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 179 GNLYDGL 186
              :||| ||||
QY 139 SDSLGYDGL 146

RESULT 10
ENTRY  #T03947 #type complete
TITLE  adenosylmethionine decarboxylase (EC 4.1.1.50) - maize
ALTERNATE_NAMES S-adenosylmethionine decarboxylase
ORGANISM #formal_name Zea mays #common_name maize
DATE    23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
              23-Apr-1999
ACCESSIONS T03947
REFERENCE   Z15155
#authors    Michael, A.J.
#submission submitted to the EMBL Data Library, September 1996
#accession  T03947
#status     preliminary; translated from GB/EMBL/DDBJ
#molecule_type mRNA
#residues   1-400 #label MIC
#cross-references EMBL:Y07767; NID:ei004964; PID:e265487
KEYWORDS    carbon-carbon lyase; carboxy-lyase
SUMMARY     #length 400 #molecular-weight 43515 #checksum 3050
Query Match   85.5%; Score 47; DB 2; Length 400;
Best Local Similarity 75.0%; Pred. No. 1.61e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 276 TALSYDGL 283
              :||| ||||
QY 139 SDSLGYDGL 146

```



```

RESULT 11
ENTRY YGPLV3 #type complete
TITLE alpha-aminoacyl-cysteine-lyase synthetase (EC 6.2.1.1) -
        Penicillium chrysogenum (strain Olil3)
ALTERNATE_NAMES ACV synthetase
ORGANISM #formal_name Penicillium chrysogenum
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
        18-Jun-1999
ACCESSIONS S13134
REFERENCE Smith, D.J.; Earl, A.J.; Turner, G.
#authors EMO J. (1990) 9:2743-2750
#journal The multifunctional peptide synthetase performing the first
#title step of penicillin biosynthesis in Penicillium chrysogenum
        is a 421 073 dalton protein similar to Bacillus brevis
        peptide antibiotic synthetases.
#cross-references MUID:90360984
#accession S13134
#molecule_type DNA
#residues 1-3746 #label SMI
#cross-references GB:X54296; NID:g3117; PIDN:CAA38195.1; PID:g3118
COMMENT This multifunctional enzyme catalyzes the formation of
        alpha-aminoacyl-cysteine-lyase by activating the three
        precursor amino acids in the L form, racemizing the L to
        D-valine, and carrying out the polymerization steps to form the
        tripeptide, which is the first common intermediate of penicillins
        and cephalosporins.
GENETICS
#gene pcbAB
CLASSIFICATION #superfamily alpha-aminoacyl-cysteine-lyase synthetase;
        acetate--CoA ligase homology; acyl carrier protein
        homology; gramicidin S synthetase I repeat homology
KEYWORDS cephalosporin biosynthesis; duplication; ligase; penicillin
        biosynthesis; phosphopantetheine; phosphoprotein
FEATURE 278-891 #domain gramicidin S synthetase I repeat homology #label
        GRS1\
        338-804 #domain acetate--CoA ligase homology #label ACL1\
        821-891 #domain acyl carrier protein homology #label ACP1\
        1372-1975 #domain gramicidin S synthetase I repeat homology #label
        GRS2\
        1432-1889 #domain acetate--CoA ligase homology #label ACL2\
        1905-1975 #domain acyl carrier protein homology #label ACP2\
        2454-3062 #domain gramicidin S synthetase I repeat homology #label
        GRS3\
        2515-2977 #domain acetate--CoA ligase homology #label ACL3\
        2994-3062 #domain acyl carrier protein homology #label ACP3\
        855,1939,3026 #binding_site phosphopantetheine (Ser) (covalent)
        #status predicted
SUMMARY #length 3746 #molecular-weight 421073 #checksum 9188
Query Match 85.5%; Score 47; DB 1; Length 3746;
Best Local Similarity 57.1%; Pred. No. 1.61e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 317 ELTYGEL 323
QY 140 DLSYGDL 146
:|:|:|:|
RESULT 12
ENTRY YGPLV8 #type complete
TITLE alpha-aminoacyl-cysteine-lyase synthetase (EC 6.2.1.1) -
        Penicillium chrysogenum (strain AS-P-78)
ALTERNATE_NAMES ACV synthetase
ORGANISM #formal_name Penicillium chrysogenum
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
        18-Jun-1999
ACCESSIONS A37886
REFERENCE Diez, B.; Gutierrez, S.; Barredo, J.L.; van Solingen, P.; van
        der Voort, L.H.M.; Martin, J.F.
#authors
GENETICS
#gene ycf39
CLASSIFICATION #superfamily alpha-aminoacyl-cysteine-lyase synthetase;
        acetate--CoA ligase homology; acyl carrier protein
        homology; gramicidin S synthetase I repeat homology
KEYWORDS cephalosporin biosynthesis; duplication; ligase; penicillin
        biosynthesis; phosphopantetheine; phosphoprotein
FEATURE 308-923 #domain gramicidin S synthetase I repeat homology #label
        GRS1\
        368-834 #domain acetate--CoA ligase homology #label ACL1\
        851-923 #domain acyl carrier protein homology #label ACP1\
        1404-2007 #domain gramicidin S synthetase I repeat homology #label
        GRS2\
        1464-1921 #domain acetate--CoA ligase homology #label ACL2\
        1937-2007 #domain acyl carrier protein homology #label ACP2\
        2486-3094 #domain gramicidin S synthetase I repeat homology #label
        GRS3\
        2547-3009 #domain acetate--CoA ligase homology #label ACL3\
        3026-3094 #domain acyl carrier protein homology #label ACP3\
        885,1971,3058 #binding_site phosphopantetheine (Ser) (covalent)
        #status predicted
SUMMARY #length 3791 #molecular-weight 425918 #checksum 3817
Query Match 85.5%; Score 47; DB 1; Length 3791;
Best Local Similarity 57.1%; Pred. No. 1.61e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 347 ELTYGEL 353
QY 140 DLSYGDL 146
:|:|:|:|
RESULT 13
ENTRY S73159 #type complete
TITLE hypothetical protein 39 - red alga (Porphyra purpurea)
        chloroplast
ORGANISM #formal_name chloroplast Porphyra purpurea
DATE 19-Mar-1997 #sequence_revision 09-May-1997 #text_change
        10-Sep-1997
ACCESSIONS S73159
REFERENCE Reith, M.; Munholland, J.
#authors Plant Mol. Biol. Rep. (1995) 13:333-335
#journal Complete nucleotide sequence of the Porphyra purpurea
#title chloroplast genome.
#accession S73159
#status preliminary; nucleic acid sequence not shown;
        translation not shown
#molecule_type DNA
#residues 1-319 #label REI
#cross-references EMBL:U38804; NID:g1276652; PID:g1276704
#note the nucleotide sequence was submitted to the EMBL Data
        Library, October 1995
GENETICS
#gene ycf39

```

#genome chloroplast
KEYWORDS chloroplast
SUMMARY #length 319 #molecular-weight 35675 #checksum 5409

Query Match 83.6%; Score 46; DB 2; Length 319;
Best Local Similarity 62.5%; Pred. No. 2.58e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 45 AELVYGD 52
QY 139 SLSYGD 146

RESULT 14
ENTRY T04099 #type complete
TITLE probable adenosymethionine decarboxylase (EC 4.1.1.50) -
ORGANISM rice
DATE 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
ACCESSION T04099
REFERENCE Z15210
#authors Michael, A.J.
#submission submitted to the EMBL Data Library, September 1996
#description Spermidine biosynthesis in rice.
#accession T04099
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-398 #label MIC
#cross-references EMBL:Y07766; NID:el004963; PID:e265442
#experimental_source cv. Nipponbare
KEYWORDS carbon-carbon lyase; carboxy-lyase
SUMMARY #length 398 #molecular-weight 43157 #checksum 7711

Query Match 83.6%; Score 46; DB 2; Length 398;
Best Local Similarity 75.0%; Pred. No. 2.58e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 276 STLAYGD 283
QY 139 SLSYGD 146

RESULT 15
ENTRY S13720 #type fragment
TITLE coat protein - arabis mosaic virus (fragment)
ORGANISM #formal_name arabis mosaic virus
DATE 19-Mar-1997 #sequence_revision 24-Jul-1997 #text_change
ACCESSION S13720
REFERENCE S13720; S77995
#authors Steinkellner, H.; Himmeler, G.; Mattanovich, D.; Katinger, H.
#journal Nucleic Acids Res. (1990) 18:7182
#title Nucleotide sequence of AMV-capsid protein-gene.
#cross-references MUID:91088343
#accession S13720
#molecule_type genomic RNA
#residues 1-506 #label STE
#cross-references EMBL:X55460
#accession S77995
#molecule_type protein
#residues 2-21 #label HIM
KEYWORDS polyprotein
FEATURE 2-506
SUMMARY #product coat protein #status predicted #label COA
#length 506 #checksum 6836

Query Match 83.6%; Score 46; DB 2; Length 506;
Best Local Similarity 71.4%; Pred. No. 2.58e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 108 DLDYGE 114
QY 139 SLSYGD 146

QY 140 DLSYGD 146

Search completed: Wed May 10 13:15:09 2000
Job time : 9 secs.

[W][A][T][E][R][M][A][N]
[S][E][R][I][E][S]
[P][R][O][T][E][I][N]
[S][E][Q][U][E][N][C][E]
[I][N]
[T][E][R][M]

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 13:08:46 2000; MasPar time 89.62 Seconds
Tabular output not generated. 2.719 Million cell updates/sec

Title: >US-09-376-430-2
Description: (139-146) from US09376430A.pep (16 of 25)
Perfect score: 55
Sequence: 1 SDLSYGD L 8

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 23.262; Variance 24.507; scale 0.949

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.		
1	49	89.1	417	1	YCE4_YEAST	HYPOTHETICAL 47.2 KD P	2.05e+00		
2	48	87.3	506	1	ARSA_MOUSE	ARYLSULFATASE A PRECUR	3.48e+00		
3	48	87.3	507	1	ARSA_HUMAN	ARYLSULFATASE A PRECUR	3.48e+00		
4	48	87.3	530	1	YBIT_ECOLI	HYPOTHETICAL ABC TRANS	3.48e+00		
5	48	87.3	989	1	YD30_YEAST	HYPOTHETICAL 112.2 KD	3.48e+00		
6	47	85.5	400	1	DCAM_MAIZE	S-ADENOSYLMETHIONINE D	5.86e+00		
7	47	85.5	3746	1	ACVS_PENCH	DELTA-(L-ALPHA-AMINOAC	5.86e+00		
8	47	85.5	3791	1	ACVT_PENCH	DELTA-(L-ALPHA-AMINOAC	5.86e+00		
9	46	83.6	319	1	YC39_PORPU	HYPOTHETICAL 35.7 KD P	9.79e+00		
10	46	83.6	398	1	DCAM_ORISA	S-ADENOSYLMETHIONINE D	9.79e+00		
11	46	83.6	506	1	POL2_ARMV	RNA2 POLYPROTEIN (150	9.79e+00		
12	45	81.8	266	1	RNC_BACSU	RIBONUCLEASE III (EC 3	1.62e+01		
13	45	81.8	319	1	YC39_ODOSI	HYPOTHETICAL 36.3 KD P	1.62e+01		
14	45	81.8	537	1	CH61_MYCLE	60 KD CHAPERONIN 1 (PR	1.62e+01		
15	45	81.8	539	1	CH61_MYCTU	60 KD CHAPERONIN 1 (PR	1.62e+01		
16	45	81.8	757	1	PIGR_BOVIN	POLYMERIC-IMMUNOGLOBUL	1.62e+01		
17	44	80.0	102	1	PIGR_PIG	POLYMERIC-IMMUNOGLOBUL	2.66e+01		
18	44	80.0	149	1	Y211_MYCPN	HYPOTHETICAL PROTEIN M	2.66e+01		
19	44	80.0	302	1	YD75_HAEIN	HYPOTHETICAL PROTEIN H	2.66e+01		
20	44	80.0	331	1	Y119_SYNY3	HYPOTHETICAL 35.3 KD P	2.66e+01		
21	44	80.0	757	1	HT16_HYDAT	TYROSINE-PROTEIN KINAS	2.66e+01		
22	44	80.0	956	1	HELL_HCMVA	PROBABLE HELICASE.	2.66e+01		
23	44	80.0	1242	1	IRSI_HUMAN	INSULIN RECEPTOR SUBST	2.66e+01		

24	44	80.0	2026	1	CYAA_YEAST	ADENYLATE CYCLASE (EC	2.66e+01
25	44	80.0	4466	1	DYHC_ANTOR	DYNEIN BETA CHAIN, CIL	2.66e+01
26	43	78.2	314	1	YC39_GUITH	HYPOTHETICAL 35.9 KD P	4.31e+01
27	43	78.2	342	1	MURB_SALTY	UDP-N-ACETYLENOLPYRUVO	4.31e+01
28	43	78.2	342	1	MURB_ECOLI	UDP-N-ACETYLENOLPYRUVO	4.31e+01
29	43	78.2	358	1	MYH2_HAEPF	MODIFICATION METHYLASE	4.31e+01
30	43	78.2	378	1	YR84_MYCTU	HYPOTHETICAL 41.1 KD P	4.31e+01
31	43	78.2	650	1	VE1_HPV72	REPLICATION PROTEIN E1	4.31e+01
32	43	78.2	655	1	FKHR_HUMAN	FORK HEAD DOMAIN PROTE	4.31e+01
33	43	78.2	697	1	NANB_STRPN	SIALIDASE B PRECURSOR	4.31e+01
34	43	78.2	1061	1	DPOL_ADE12	DNA POLYMERASE (EC 2.7	4.31e+01
35	43	78.2	3770	1	ACVS_EMENI	DELTA-(L-ALPHA-AMINOAC	4.31e+01
36	42	76.4	207	1	Y0Y2_YEAST	HYPOTHETICAL 23.2 KD P	6.94e+01
37	42	76.4	218	1	GTS_ANOGA	GLUTATHIONE S-TRANSFER	6.94e+01
38	42	76.4	275	1	YT57_MYCTU	HYPOTHETICAL 31.0 KD P	6.94e+01
39	42	76.4	277	1	YK3_YEAST	HYPOTHETICAL 32.3 KD P	6.94e+01
40	42	76.4	312	1	MDH_VIBS5	MALATE DEHYDROGENASE (6.94e+01
41	42	76.4	454	1	ATY_RAT	TYROSINE AMINOTRANSFER	6.94e+01
42	42	76.4	454	1	ATY_HUMAN	TYROSINE AMINOTRANSFER	6.94e+01
43	42	76.4	701	1	YIA5_YEAST	PUTATIVE DISULFIDE ISO	6.94e+01
44	42	76.4	857	1	V2A_CMVKT	PROBABLE RNA-DIRECTED	6.94e+01
45	42	76.4	1115	1	FRPA_NEIME	IRON-REGULATED PROTEIN	6.94e+01

ALIGNMENTS

RESULT 1
ID YCE4_YEAST STANDARD; PRT; 417 AA.
AC P2573;
DT 01-MAY-1992 (Rel. 22, Crehted)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 47.2 KD PROTEIN IN PDII 5'REGION.
GN YCL044C OR YCL44C OR YCL314.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92397595.
RA Scherens B., Messenguy F., Gigot D., Dubois E.;
RT "The complete sequence of a 9,543 bp segment on the left arm of
RT chromosome III reveals five open reading frames including glucokinase
RT and the protein disulfide isomerase.";
RL yeast 8:577-586(1992).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X59720; CAA2372.1;
DR PIR; S19373; S19373.
DR PIR; S25348; S25348.
KW Hypothetical protein.
SQ SEQUENCE 417 AA; 47156 MW; 390558E8ECF6C3B9B CRC64;
Query Match 89.1%; Score 49; DB 1; Length 417;
Best Local Similarity 75.0%; Pred. No. 2.05e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 93 SDLSFGEL 100
Qy 139 SDLSYGD L 146
RESULT 2
ID ARSA_MOUSE STANDARD; PRT; 506 AA.
AC P50428;
DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ARYLSULFATASE A PRECURSOR (EC 3.1.6.8) (ASA) (CEREBROSIDE-SULFATASE).
 GN ARSA OR AS2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SV, AND C57BL/6J;
 RX MEDLINE; 94245194.
 RA Kreysing J., Polten A., Hess B., von Figura K., Menz K., Steiner F.,
 RA Gieselmann V.,
 RT "Structure of the mouse arylsulfatase A gene and cDNA."
 RL Genomics 19:249-256(1994).
 RN [2]
 RP SEQUENCE OF 32-66 FROM N.A.
 RX MEDLINE; 92241876.
 RA Grompe M., Pieretti M., Caskey C.T., Ballabio A.;
 RT "The sulfatase gene family: cross-species PCR cloning using the MOPAC
 RT technique."
 RL Genomics 12:755-760(1992).
 CC -1- FUNCTION: HYDROLYSES CEREBROSIDE SULFATE.
 CC -1- CATALYTIC ACTIVITY: A CEREBROSIDE 3-SULFATE + H(2)O -> A
 CC CEREBROSIDE + SULFATE.
 CC -1- SUBUNIT: EXISTS BOTH AS A SINGLE CHAIN OF 58 KD (COMPONENT A)
 CC OR AS A CHAIN OF 50 KD (COMPONENT B) LINKED BY DISULFIDE BOND(S)
 CC TO A 7 KD CHAIN (COMPONENT C).
 CC -1- SUBCELLULAR LOCATION: LYOSOMAL.
 CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X73231; CAA51703.1; -
 DR EMBL; X73230; CAA51702.1; -
 DR EMBL; M82876; AAA37260.1; -
 DR HSP; P15289; IADK.
 DR MGD; MGI:88077; AS2.
 DR PROSITE; PS00523; SULFATASE_1; 1.
 DR PROSITE; PS00149; SULFATASE_2; 1.
 DR PRAM; PF00884; Sulfatase; 1.
 KW Hydrolase; Signal; Glycoprotein; Lysosome.
 FT SIGNAL 1 17
 FT CHAIN 18 506
 FT MOD_RES 68
 FT
 FT
 FT DISULFID 155 171
 FT BY SIMILARITY.
 FT DISULFID 160 167
 FT BY SIMILARITY.
 FT DISULFID 299 413
 FT BY SIMILARITY.
 FT DISULFID 487 499
 FT BY SIMILARITY.
 FT DISULFID 488 501
 FT BY SIMILARITY.
 FT DISULFID 492 498
 FT BY SIMILARITY.
 FT CARBOHYD 157 157
 FT POTENTIAL.
 FT CARBOHYD 183 183
 FT POTENTIAL.
 FT CARBOHYD 349 349
 FT SEQUENCE 506 AA; 53776 MW; 74BB16401CF18DF6 CRC64;
 SQ
 Query Match 87.3%; Score 48; DB 1; Length 506;
 Best Local Similarity 85.7%; Pred. No. 3.48e+00;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 29 DLGIGDL 35
 QY 140 DLSYDGL 146
 RESULT 3

ID ARSA_HUMAN STANDARD; PRT; 507 AA.
 AC P15289;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE ARYLSULFATASE A PRECURSOR (EC 3.1.6.8) (ASA) (CEREBROSIDE-SULFATASE).
 GN ARSA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 90361046.
 RA Kreysing J., von Figura K., Gieselmann V.;
 RT "Structure of the arylsulfatase A gene."
 RL Eur. J. Biochem. 191:627-631(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 89093115.
 RA Stein C., Gieselmann V., Kreysing J., Schmidt B., Pohlmann R.,
 RA Waheed A., Meyer H.E., O'Brien J.S., von Figura K.;
 RT "Cloning and expression of human arylsulfatase A."
 RL J. Biol. Chem. 264:1252-1259(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Kerlavage A.R., Fuldner R.A., Phillips C.A., Venter J.C.;
 RT Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 19-33 AND 434-479, AND SUBUNITS.
 RX MEDLINE; 92338230.
 RA Fujii T., Kobayashi T., Honke K., Gasa S., Ishikawa M., Shimizu T.,
 RA Makita A.;
 RT "Proteolytic processing of human lysosomal arylsulfatase A."
 RL Biochim. Biophys. Acta 1122:93-98(1992).
 RN [5]
 RP PARTIAL SEQUENCE, AND 2-AMINO-3-OXOPROPIONIC ACID MODIFICATION SITE.
 RX MEDLINE; 95354208.
 RA Schmidt B., Selmer T., Ingendoh A., von Figura K.;
 RT "A novel amino acid modification in sulfatases that is defective in
 RL multiple sulfatase deficiency."
 RL Cell 82:271-278(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
 RX MEDLINE; 98191325.
 RA Lukatela G., Krauss N., Theis K., Selmer T., Gieselmann V.,
 RA von Figura K., Saenger W.;
 RT "Crystal structure of human arylsulfatase A: the aldehyde function
 RT and the metal ion at the active site suggest a novel mechanism for
 RL sulfate ester hydrolysis."
 RL Biochemistry 37:3654-3664(1998).
 RN [7]
 RP REVIEW ON MLD VARIANTS.
 RX MEDLINE; 95170731.
 RA Gieselmann V., Zlotogora J., Harris A., Wenger D.A., Morris C.P.;
 RT "Molecular genetics of metachromatic leukodystrophy."
 RL Hum. Mutat. 4:233-242(1994).
 RN [8]
 RP VARIANT MLD GLN-84.
 RX MEDLINE; 92344341.
 RA Kappler J., von Figura K., Gieselmann V.;
 RT "Late-onset metachromatic leukodystrophy: molecular pathology in two
 RL siblings."
 RL Ann. Neurol. 31:256-261(1992).
 RN [9]
 RP VARIANT MLD PHE-96.
 RX MEDLINE; 91328147.
 RA Gieselmann V., Fluharty A.L., Tonnesen T., von Figura K.;
 RT "Mutations in the arylsulfatase A pseudodeficiency allele causing
 RL metachromatic leukodystrophy."
 RL Am. J. Hum. Genet. 49:407-413(1991).
 RN [10]
 RP VARIANT MLD ASP-99.
 RX MEDLINE; 91206410.

RA Kondo R., Wakamatsu N., Yoshino H., Fukuhara N., Miyatake T.,
 RA Tsuji S.;
 RA "Identification of a mutation in the arylsulfatase A gene of a
 RT patient with adult-type metachromatic leukodystrophy.";
 RL Am. J. Hum. Genet. 48:971-978(1991).
 RN [11]
 RP VARIANT MLD LEU-426, AND VARIANTS CYS-193 AND SER-391.
 RX MEDLINE; 91074201.
 RA Polten A., Fluhrhart A.L., Fluhrhart C.B., Kappler J., von Figura K.,
 RA Gieselmann V.;
 RA "Molecular basis of different forms of metachromatic leukodystrophy.";
 RT New Engl. J. Med. 324:18-22(1991).
 RL [12]
 RN VARIANT MLD SER-122.
 RX MEDLINE; 94063853.
 RA Honke K., Kobayashi T., Fujii T., Gasa S., Xu M., Takamaru Y.,
 RA Kondo R., Tsuji S., Makita A.;
 RA "An adult-type metachromatic leukodystrophy caused by substitution of
 RT serine for glycine-122 in arylsulfatase A.";
 RL Hum. Genet. 92:451-456(1993).
 RN [13]
 RP VARIANT MLD ARG-245.
 RX MEDLINE; 93319832.
 RA Hasegawa Y., Kawane H., Eto Y.;
 RA "Mutations in the arylsulfatase A gene of Japanese patients with
 RT metachromatic leukodystrophy.";
 RL DNA Cell Biol. 12:493-498(1993).
 RN [15]
 RP VARIANT MLD MET-274.
 RX MEDLINE; 94004907.
 RA Harvey J.S., Nelson P.V., Carey W.F., Robertson E.F., Morris C.P.;
 RA "An arylsulfatase A (ARSA) missense mutation (T274M) causing late-
 RT infantile metachromatic leukodystrophy.";
 RL Hum. Mutat. 2:261-267(1993).
 RN [16]
 RP VARIANT MLD SER-309.
 RX MEDLINE; 93318834.
 RA Kreysing J., Bohne W., Bosenberg C., Marchesini S., Turpin J.C.,
 RA Baumann N., von Figura K., Gieselmann V.;
 RA "High residual arylsulfatase A (ARSA) activity in a patient with
 RT late-infantile metachromatic leukodystrophy.";
 RL Am. J. Hum. Genet. 53:339-346(1993).
 RN [17]
 RP VARIANT SER-350.
 RX MEDLINE; 90083282.
 RA Gieselmann V., Polten A., Kreysing J., von Figura K.;
 RA "Arylsulfatase A pseudodeficiency: loss of a polyadenylation signal
 RT and N-glycosylation site.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:9436-9440(1989).
 RN [18]
 RP VARIANT MLD LEU-426.
 RX MEDLINE; 93202658.
 RA Barth M.L., Fensom A., Harris A.;
 RA "Prevalence of common mutations in the arylsulfatase A gene in
 RT metachromatic leukodystrophy patients diagnosed in Britain.";
 RL Hum. Genet. 91:73-77(1993).
 RN [19]
 RP VARIANTS MLD VAL-212; VAL-224 AND TYR-295.
 RX MEDLINE; 94154687.
 RA Barth M.L., Fensom A., Harris A.;
 RA "Missense mutations in the arylsulfatase A genes of metachromatic
 RT leukodystrophy patients.";
 RL Hum. Mol. Genet. 2:2117-2121(1993).
 RN [20]
 RP VARIANTS MLD LEU-82; TYR-172; CYS-201; GLN-311; VAL-335 AND TRP-390.
 RX MEDLINE; 96047150.

RA Barth M.L., Fensom A., Harris A.;
 RA "Identification of seven novel mutations associated with
 RT metachromatic leukodystrophy.";
 RL Hum. Mutat. 5:170-176(1995).
 RN [21]
 RP CHARACTERIZATION OF VARIANTS MET-274 AND VAL-335.
 RX MEDLINE; 96303701.
 RA Hess B., Kafert S., Heinisch U., Wenger D.A., Zlotogora J.,
 RA Gieselmann V.;
 RA "Characterization of two arylsulfatase A missense mutations D335V and
 RT T274M causing late infantile metachromatic leukodystrophy.";
 RL Hum. Mutat. 7:311-317(1996).
 RN [22]
 RP VARIANTS MLD.
 RX MEDLINE; 97245886.
 RA Draghia R., Letourneur F., Dragan C., Manicom J., Blanchot C.,
 RA Kahn A., Poenaru L., Caillaud C.;
 RA "Metachromatic leukodystrophy: identification of the first deletion in
 RT exon 1 and of nine novel point mutations in the arylsulfatase A
 RL gene.";
 RL Hum. Mutat. 9:234-242(1997).
 RN [23]
 RP VARIANT MLD SER-406--THR-408 DEL.
 RX MEDLINE; 98141126.
 RA Regis S., Filocamo M., Stroppiano M., Corsolini F., Caroli F.,
 RA Gatti R.;
 RA "A 9-bp deletion (2320del9) on the background of the arylsulfatase A
 RT pseudodeficiency allele in a metachromatic leukodystrophy patient and
 RT in a patient with nonprogressive neurological symptoms.";
 RL Hum. Genet. 102:50-53(1998).
 RN [24]
 RP VARIANTS MLD LEU-135 AND SER-179.
 RX MEDLINE; 98260871.
 RA Gomez-Lira M., Perusi C., Mottes M., Pignatti P.F., Manfredi M.,
 RA Rizzuto N., Salviati A.;
 RA "Molecular genetic characterization of two metachromatic
 RT leukodystrophy patients who carry the T799G mutation and show
 RT different phenotypes; description of a novel null-type mutation.";
 RL Hum. Genet. 102:459-463(1998).
 RN [25]
 RP VARIANTS MLD GLN-390 AND TYR-397.
 RX MEDLINE; 98112481.
 RA Coulter-Mackie M.B., Gagnier L.;
 RA "Two novel mutations in the arylsulfatase A gene associated with
 RT juvenile (R390Q) and adult onset (H397Y) metachromatic
 RT leukodystrophy.";
 RL Hum. Mutat. Suppl. 1:S254-S256(1998).
 RN [26]
 RP VARIANT HIS-496.
 RX MEDLINE; 98415722.
 RA Ricketts M.H., Poretz R.D., Manowitz P.;
 RA "The R496H mutation of arylsulfatase A does not cause metachromatic
 RT leukodystrophy.";
 RL Hum. Mutat. 12:238-239(1998).
 RN [27]
 RP Note: remainder of annotations omitted.

Query Match 87.3%; Score 48; DB 1; Length 507;
 Best Local Similarity 85.7%; Pred. No. 3.48e+00;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 30 DLGYGDL 36
 11:1111
 QY 140 DLGYGDL 146

RESULT 4
 ID YBIT_ECOLI STANDARD; PRT; 530 AA.
 AC P75790;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBIT.

GN YBIT.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.,
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE: 97061202.
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.,
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map";
RL DNA Res. 3:137-155(1996).
[1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
(ABC TRANSPORTERS). EF-3 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

DR EMBL; AE000184; AAC73907.1; -;
DR EMBL; D90719; BAA35501.1; -;
DR EMBL; D90720; BAA35508.1; -;
DR ECOGENE; EG13325; YBIT.
DR PROSITE; PS00211; ABC_TRANSPORTER; FALSE_NEG.
DR PFAM; PF00005; ABC_tran; 2.
KW Hypothetical protein; ATP-binding; Transport; Repeat.
FT NP_BIND 34 41
FT NP_BIND 352 359 ATP (POTENTIAL).
SQ SEQUENCE 530 AA; 59858 MW; EE75D66B5A58729A CRC64;
Query Match 87.3%; Score 48; DB 1; Length 530;
Best Local Similarity 62.5%; Pred. No. 3.48e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 223 ADLIDYGEI 230
Qy 139 SDSLSDGL 146
:|||||:

RESULT 5
ID YD30_YEAST STANDARD; PRT; 989 AA.
AC P32898; Q04068;
DT 01-OCT-1993 (Rel. 27, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 112.2 KD PROTEIN IN TIF35-NPL3 INTERGENIC REGION (ORF1).
GN YD430C OR D9461.18
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomycetes.
[1]
RN SEQUENCE FROM N.A.
DR Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E., Berno A.,
RA Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,

RA Hunnicke-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,
RA Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,
RA Winant A., Yelton M., Botstein D., Davis R.W.,
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE OF 11-166 FROM N.A.
RX MEDLINE: 94040759.
RA Ellis E.M., Reid G.A.,
RT "The Saccharomyces cerevisiae MTS1 gene encodes a putative
RNA-binding protein involved in mitochondrial protein targeting.";
RL Gene 132:175-183(1993).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

DR EMBL; X70951; CAA50290.1; -;
DR EMBL; U33007; AAB54877.1; -;
DR FIR; S31321; S31321.
KW Hypothetical protein.
FT CONFLICT 11 18 YAAQAQAVR -> MPKHPRLA (IN REF. 2).
SQ SEQUENCE 989 AA; 112180 MW; 6CAD2EE5A6080D9C CRC64;
Query Match 87.3%; Score 48; DB 1; Length 989;
Best Local Similarity 75.0%; Pred. No. 3.48e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 220 TDLRYGDM 227
Qy 139 SDSLSDGL 146
:|||||:

RESULT 6
ID DCAM_MAIZE STANDARD; PRT; 400 AA.
AC Q24575;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE S-ADENOSYLMETHIONINE DECARBOXYLASE PROENZYME (EC 4.1.1.50) (ADOMETDC)
DE (SAMDC).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Zea.
[1]
RN SEQUENCE FROM N.A.
RA Michael A.J.,
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: S-ADENOSYLMETHIONINE -> (S-DEOXY-5-ADENOSYL)
(3-AMINOPROPYL)METHYLSULFONIUM SALT + CO(2).
CC -!- COFACTOR: REQUIRES A PYRUVYL GROUP FOR ITS ACTIVITY.
CC -!- PATHWAY: DECARBOXYLATION OF S-ADENOSYLMETHIONINE PROVIDES THE
AMINOPROPYL MOIETY REQUIRED FOR SPERMIDINE AND SPERMINE
BIOSYNTHESIS FROM PUTRESCINE.
CC -!- SIMILARITY: BELONGS TO THE ADOMETC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

DR EMBL; Y07767; CAA69075.1; -;
DR PROSITE; PS01336; ADOMETDC; 1.
DR PFAM; PF01536; SAM_decarbox; 1.
KW Spermidine biosynthesis; Lyase; Decarboxylase; Pyruvate; Zymogen.

CC EMBL; X54296; CAA38195.1; -.
 DR PIR: S13134; YGPLV3.
 DR HSSP: P14687; IAMU.
 DR PRINTS: PRO0154; AMPBINDING.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
 DR PROSITE; PS00455; AMP_BINDING; 3.
 DR PROSITE; PS50075; ACP_DOMAIN; 3.
 DR PFAM: PF00501; AMP-binding; 3.

CC EMBL; M57425; AAA63415.1; -
DR PIR; A37886; YGPLV8.
DR HSSP; P14687; 1AMU.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHANTHINE; 3.
DR PROSITE; PS00435; AMP BINDING; 3.
CC

DR PROSITE: PS00075; ACP_DOMAIN; 3.
 DR PFAM: PF00501; AMP-binding; 3.
 DR PFAM: PF00550; PP-binding; 3.
 DR PFAM: PF00668; DUF4; 3.
 DR PFAM: PF00975; Thioesterase; 1.
 DR KW Ligase: Antibiotic biosynthesis; Multifunctional enzyme;
 DR KW Repeat: Phosphopantetheine.
 FT REPEAT 307 1068
 FT REPEAT 1392 2157 DOMAIN 1 (ADIPATE-ACTIVATING).
 FT REPEAT 2574 3242 DOMAIN 2 (CYSTEINE-ACTIVATING).
 FT DOMAIN 853 924 DOMAIN 3 (VALINE-ACTIVATING).
 FT DOMAIN 1939 2008 ACYL CARRIER (ACP).
 FT DOMAIN 3028 3095 ACYL CARRIER (ACP).
 FT BINDING 885 885 ACYL CARRIER (ACP).
 FT BINDING 1971 1971 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT BINDING 3058 3058 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT ACT_SITE 3631 3631 THIOESTERASE (BY SIMILARITY).
 SQ SEQUENCE 3791 AA; 425914 MW; 1B628000F8A00492 CRC64;
 Query Match 85.5%; Score 47; DB 1; Length 3791;
 Best Local Similarity 57.1%; Pred. No. 5.86e+00;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Db 347 ELYGEL 353
 QY 140 DLSYGD 146
 RESULT 9
 ID YC39_PORPU STANDARD; PRT; 319 AA.
 AC P51238;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHETICAL 33.7 KD PROTEIN YCF39 (ORF319).
 GN YCF39.
 OS Porphyra purpurea.
 OC Chloroplast.
 OG Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
 RN [1]
 RC STRAIN-AVONPORT;
 RA Reith M.E., Munholland J.;
 RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast genome."
 RL Plant Mol. Biol. Rep. 13:333-335(1995).
 CC -!- SIMILARITY: BELONGS TO THE YCF39 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U38804; AAC08124.1;
 CC MENDEL; 17196; PORPU; ycf39; mml7196.
 CC KW Chloroplast; Hypothetical protein.
 SQ SEQUENCE 319 AA; 35675 MW; 8F266C933BB10E8D CRC64;
 Query Match 83.6%; Score 46; DB 1; Length 319;
 Best Local Similarity 62.5%; Pred. No. 9.79e+00;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 45 AELVYGD 52
 QY 139 SDLSYGD 146
 RESULT 10
 ID DCAM_ORYSA STANDARD; PRT; 398 AA.
 AC O24215; O81269;

DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE S-ADENOSYLMETHIONINE DECARBOXYLASE PROENZYME (EC 4.1.1.50) (ADOMETDC)
 DE (SAMDC).
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 OC Poaceae; Oryza.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. NIPPONBARE;
 RA Michael A.J.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Li Z.Y., Chen S.Y.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: S-ADENOSYLMETHIONINE -> (5-DEOXY-5-ADENOSYL)
 CC (3-AMINOPROPYL)METHYLSULFONIUM SALT + CO(2).
 CC -!- COFACTOR: REQUIRES A PYRUVYL GROUP FOR ITS ACTIVITY.
 CC -!- PATHWAY: DECARBOXYLATION OF S-ADENOSYLMETHIONINE PROVIDES THE
 CC AMINOPROPYL MOIETY REQUIRED FOR SPERMIDINE AND SPERMINE
 CC BIOSYNTHESIS FROM PUTRESCINE.
 CC -!- SIMILARITY: BELONGS TO THE ADOMETC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Y07766; CAA69074.1;
 CC DR EMBL; AF067194; AAC79990.1;
 CC DR PROSITE; PS01336; ADOMETDC; 1.
 CC DR PFAM; PF01536; SAM_decarbox; 1.
 CC KW Spermidine biosynthesis; Lyase; Decarboxylase; Pyruvate; Zymogen.
 FT FT CHAIN 1 77 S-ADENOSYLMETHIONINE DECARBOXYLASE BETA
 FT CHAIN 78 398 CHAIN (BY SIMILARITY).
 FT SITE 77 78 S-ADENOSYLMETHIONINE DECARBOXYLASE ALPHA
 FT MOD_RES 78 78 CLEAVAGE (NONHYDROLYTICAL)
 FT ACT_SITE 18 18 (BY SIMILARITY).
 FT ACT_SITE 21 21 CONVERTED TO A PYRUVYL GROUP
 FT ACT_SITE 92 92 (BY SIMILARITY).
 FT CONFLICT 3 3 IMPORTANT FOR CATALYTIC ACTIVITY (BY
 FT CONFLICT 76 85 SIMILARITY).
 FT CONFLICT 135 135 IMPORTANT FOR CATALYTIC ACTIVITY (BY
 FT CONFLICT 250 250 SIMILARITY).
 SQ SEQUENCE 398 AA; 43309 MW; 1324EDA871B2AF39 CRC64;
 Query Match 83.6%; Score 46; DB 1; Length 398;
 Best Local Similarity 75.0%; Pred. No. 9.79e+00;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 276 STLAYGD 283
 QY 139 SDLSYGD 146
 RESULT 11
 ID POL2_ARMV STANDARD; PRT; 506 AA.
 AC P24819;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)

Thu May 11 06:50:01 2000

```

or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D64116; BAA10976.1; -
DR SUBMITTER; BG11537; RNC.
DR PROSITE; P500517; RIBONUCLEASE_III; 1.
DR PFAM; PF00035; dsrm; 1.
DR PFAM; PF00636; Ribonuclease_3; 1.
KW Hydrolase; Nuclease; Endonuclease; RNA-binding.
FT DOMAIN 226 242
SQ SEQUENCE 266 AA; 30386 MW; 467EA397D53B9060 CRC64;

Query Match 81.8%; Score 45; DB 1; Length 266;
Best Local Similarity 71.4%; Pred. No. 1.62e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 108 DLSYDGL 114
QY 140 DLSYDGL 146
:|:|:|

RESULT 13
ID YC39_ODOSI STANDARD; PRT; 319 AA.
AC P49534;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 36.3 KD PROTEIN YCF39 (ORF319).
GN YCF39
OS Odontella sinensis.
OG Chloroplast.
OC Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC Bidulphiophycidae; Eupodiscaceae; Eupodiscaceae; Odontella.
RN [1]
RA Kowalik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U.;
RT "The chloroplast genome of a chlorophyll a+c-containing alga,
RL Odontella sinensis.";
RL Plant Mol. Biol. Rep. 13:336-342(1995).
CC -!- SIMILARITY: BELONGS TO THE YCF39 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z67753; CAA91705.1; -
DR MENDEL; 5324; ODSi; ycf39; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 319 AA; 36343 MW; 15B974DFED2BEE09 CRC64;

Query Match 81.8%; Score 45; DB 1; Length 319;
Best Local Similarity 62.5%; Pred. No. 1.62e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 45 AELIYDGL 52
QY 139 DLSYDGL 146
:|:|:|

RESULT 14
ID CH61_MYCLE STANDARD; PRT; 537 AA.
AC P37578;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE 60 KD CHAPERONIN 1 (PROTEIN CPN60 1) (GROEL PROTEIN 1).
GN GROEL-1 OR GROEL OR B229_C3_248 OR B1620_C3_228.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.

```

```

11-JUN-1994 (Rel. 29, Last annotation update)
DE RNA2 POLYPROTEIN (150 KD PROTEIN) [CONTAINS: COAT PROTEIN] (FRAGMENT).
OS Arabis mosaic virus (AMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC Nepovirus.
[1]
SEQUENCE FROM N.A., AND SEQUENCE OF 2-21.
RX MEDLINE; 91088343; Himmler G., Mattanovich D., Katinger H.;
RA Steinkellner H., Mattanovich D., Katinger H.;
RT "Nucleotide sequence of AMV-capsid protein-gene.";
RL Nucleic Acids Res. 18:7182-7182(1990).
CC -!- PTM: THE RNA2 POLYPROTEIN IS CLEAVED BY AN RNA1-ENCODED PROTEASE
CC TO YIELD THE MATURE COAT PROTEIN AND AN 84 KD PROTEIN WHICH IS
CC FURTHER CLEAVED INTO TWO PRODUCTS OF APPROXIMATELY 46 AND 48 KD.
CC -!- SIMILARITY: TO THE RNA2 POLYPROTEIN OF OTHER NEPOVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X55460; CAA39108.1; ALT_SEQ.
DR PUR; S13720; S13720.
KW Polyprotein; Coat protein.
FT DOMAIN 1 506
FT NON_TER 1 506
SQ SEQUENCE 506 AA; 55866 MW; 8883604ARA485DD82 CRC64;

Query Match 83.6%; Score 46; DB 1; Length 506;
Best Local Similarity 71.4%; Pred. No. 9.79e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 108 DLSYDGL 114
QY 140 DLSYDGL 146
:|:|:|

RESULT 12
ID RNC_BACSU STANDARD; PRT; 266 AA.
AC P51833;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE RIBONUCLEASE III (EC 3.1.26.3) (RNASE III).
GN RNC
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
[1]
SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE; 96257247.
RA Oguro A., Kakeshita H., Takamatsu H., Nakamura K., Yamane K.;
RT "The effect of Srb, a homologue of the mammalian SRP receptor alpha-
RT subunit, on Bacillus subtilis growth and protein translocation.";
RL Gene 172:17-24(1996).
CC -!- FUNCTION: DIGESTS DOUBLE-STRANDED RNA. INVOLVED IN THE PROCESSING
CC OF RIBOSOMAL RNA PRECURSORS AND OF SOME MRNAs (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ENDONUCLEOLYTIC CLEAVAGE TO 5'-PHOSPHO-
CC MONESTER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: CONTAINS 1 DIRM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE RNASE III FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

[1]
SEQUENCE FROM N.A.
MEDLINE; 92374850.
RA de Wit T.F.R., Bekelie S., Osland A., Miko T.L., Hermans P.W.M.,
RA van Soelingen D., Drijfhout J., Schoeningh R., Janson A.A.M.,
RA Thole J.E.R.;
RT "Mycobacteria contain two groEL genes: the second Mycobacterium
RL leorae groEL gene is arranged in an operon with groES.";
RL Mol. Microbiol. 6:1995-2007(1992).
[2]
SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RA Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z11665; NOT_ANNOTATED_CDS.
DR EMBL; U00015; AAC43228.1; -.
DR EMBL; U00020; AAA17312.1; -.
DR PIR; S25181; S25181.
DR HSSP; P06139; IGRL.
DR PRINTS; PR00298; CHAPERONIN60.
DR PROSITE; PS00296; TCOMPLEXTCP1.
DR PFAM; PF00118; cpn60_TCP1; 1.
DR Chaperone; Antigen; Multigene family.
SQ SEQUENCE 537 AA; 55816 MW; F2355B1BCFAED50F CRC64;

Query Match 81.8%; Score 45; DB 1; Length 537;
Best Local Similarity 75.0%; Pred. No. 1.62e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 480 STLGYGDL 487
QY 139 SDSLGYGDL 146
|:|:|:|:|
|:|:|:|:|

RESULT 15
ID CH61-MYCTU STANDARD; PRT; 539 AA.
AC Q59573; Q59581;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE 60 KD CHAPERONIN 1 (PROTEIN CPN60-1) (GROEL PROTEIN 1).
GN GROEL OR GROEL-1 OR RV3417C OR MTCY78.12.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
[1]
SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE; 98295987.
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the

```

```

RT complete genome sequence.";
RL Nature 393:537-544(1998).
[2]
SEQUENCE FROM N.A.
RC STRAIN=ERDMANN;
RX MEDLINE; 93219332.
RA Kong T.H., Coates A.R.M., Butcher P.D., Hickman C.J., Shinnick T.M.;
RT "Mycobacterium tuberculosis expresses two chaperonin-60 homologs.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:2608-2612(1993).
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z77165; CAB01006.1; -.
DR EMBL; X60350; CAA42909.1; -.
DR HSSP; P06139; IGRL.
DR PRINTS; PR00298; CHAPERONIN60.
DR PROSITE; PS00296; TCOMPLEXTCP1.
DR PFAM; PF00118; cpn60_TCP1; 1.
DR TUBERCULIST; RV3417C; -.
DR Chaperone; ATP-binding; Multigene family; Antigen; Cell wall.
FT VARIANT 467 467 N -> K (IN STRAIN ERDMANN).
SQ SEQUENCE 539 AA; 55877 MW; 3E0B93164C091B63 CRC64;

Query Match 81.8%; Score 45; DB 1; Length 539;
Best Local Similarity 100.0%; Pred. No. 1.62e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 483 LSYGDL 488
QY 141 LSYGDL 146
|:|:|:|:|
|:|:|:|:|

Search completed: Wed May 10 13:10:25 2000
Job time : 99 secs.

```

 M P S R C H
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.

Copyright (c) 1993-1998 University of Edinburgh, U.K.

Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 13:10:43 2000; MasPar time 228.91 Seconds
 2.423 Million cell updates/sec

Tabular output not generated.

Title: >US-09-376-430-2
 Description: (139-146) from US09376430A.pep (16 of 25)
 Perfect Score: 55

Sequence: 1 SLSYGDGL 8

Scoring table: PAM 150
 Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: sptrembl12
 1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
 5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
 9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
 13:sp_vertebrate 14:sp_virus

Statistics: Mean 22.177; Variance 26.675; scale 0.831

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	51	92.7	410	2	Q9X4V3	2.75e+00
2	50	90.9	2122	2	Q9XBC2	4.52e+00
3	50	90.9	3158	2	O52819	4.52e+00
4	49	89.1	281	1	O30017	7.40e+00
5	48	87.3	284	1	O58324	1.20e+01
6	48	87.3	430	3	O94421	1.20e+01
7	48	87.3	1276	11	P97260	1.20e+01
8	48	87.3	1277	4	Q12770	1.20e+01
9	47	85.5	377	1	O26161	1.94e+01
10	46	83.6	194	14	O66080	3.11e+01
11	46	83.6	508	2	P74942	3.11e+01
12	46	83.6	737	14	Q65028	3.11e+01
13	46	83.6	1084	14	Q65029	4.95e+01
14	45	81.8	78	1	Q9Y8M8	4.95e+01
15	45	81.8	155	10	O23368	4.95e+01
16	45	81.8	195	2	O9X6S1	4.95e+01
17	45	81.8	249	2	O31734	4.95e+01
18	45	81.8	320	2	P74206	4.95e+01
19	45	81.8	334	14	Q82541	4.95e+01
20	45	81.8	406	2	Q46377	4.95e+01

21	45	81.8	684	14	Q98218	4.95e+01
22	45	81.8	950	11	O54853	4.95e+01
23	45	81.8	1025	2	Q48153	4.95e+01
24	45	81.8	1077	2	Q92A21	4.95e+01
25	45	81.8	1195	11	O54852	4.95e+01
26	45	81.8	2105	5	Q17679	4.95e+01
27	44	80.0	251	2	O92GK3	7.82e+01
28	44	80.0	340	1	O9YEV7	7.82e+01
29	44	80.0	366	1	O58946	7.82e+01
30	44	80.0	366	1	O29817	7.82e+01
31	44	80.0	491	10	O22714	7.82e+01
32	44	80.0	532	2	Q50176	7.82e+01
33	44	80.0	539	2	O53005	7.82e+01
34	44	80.0	647	2	P73914	7.82e+01
35	44	80.0	815	3	O94598	7.82e+01
36	44	80.0	869	10	O23253	7.82e+01
37	44	80.0	955	14	O68558	7.82e+01
38	44	80.0	1251	6	Q28224	7.82e+01
39	44	80.0	1504	2	Q2GAG6	7.82e+01
40	44	80.0	1575	2	P94904	7.82e+01
41	44	80.0	2448	2	Q51338	7.82e+01
42	43	78.2	380	2	P71077	1.22e+02
43	43	78.2	509	1	O28347	1.22e+02
44	43	78.2	628	3	O42939	1.22e+02
45	43	78.2	855	5	Q23371	1.22e+02

ALIGNMENTS

RESULT 1
 ID Q9X4V3 PRELIMINARY; PRT; 410 AA.
 AC Q9X4V3;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE CPS21.
 GN CPS21.
 OS Streptococcus suis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-10
 RX MEDLINE; 99184998.
 RA SMITH H.E., DAMMAN M., VAN DER VELDE J., WAGENAR F., WISELINK H.J.,
 RA STOCKHOPE-ZURWIENEN N., SMITS M.A.;
 RT "Identification and characterization of the cps locus of Streptococcus
 RT suis serotype 2: the capsule protects against phagocytosis and is an
 RT important virulence factor".
 RL Infect. Immun. 67:1750-1756(1999).
 DR EMBL; AF118389; AAD24455.1; -;
 SQ SEQUENCE 410 AA; 46959 MW; 3AB288C7 CRC32;

Query Match 92.7%; Score 51; DB 2; Length 410;
 Best Local Similarity 75.0%; Pred. No. 2.75e+00;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 295 ADLAYGDL 302
 :||:||||
 QY 139 SLSYGDGL 146

RESULT 2
 ID Q9XBC2 PRELIMINARY; PRT; 2122 AA.
 AC Q9XBC2;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE PUTATIVE PEPTIDE SYNTHETASE (FRAGMENT).
 GN C2A382.31.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycetaceae; Streptomyces.

```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-A3(2);
RA  LENNARD N., HARRIS B.;
RT  "Sequencing and analysis of genes involved in the biosynthesis of a
RL  vancomycin group antibiotic.";
RN  Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN-A3(2);
RA  BENTLEY S.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL  Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN-A3(2);
RA  VAN WAGENINGEN A., KIRKPATRICK P., WILLIAMS D., HARRIS B., KERSHAW J.,
RT  LENNARD N., JONES M., JONES S., SOLENERG P.;
RL  "Sequencing and analysis of genes involved in the biosynthesis of a
RT  vancomycin group antibiotic.";
RL  Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR  EMBL: AL078635; CAB45052.1; -
DR  PROSITE: PS00455; AMP_BINDING; 2.
FT  NON_TER 2122 2122
SQ  SEQUENCE 2122 AA; 225386 MW; A4110B62 CRC32;

Query Match      90.9%; Score 50; DB 2; Length 2122;
Best Local Similarity 62.5%; Pred. No. 4.52e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db  1055 GELSYGEL 1062
QY  139 SDSLSDGL 146
      :::::|

RESULT  3
ID  O52819      PRELIMINARY;      PRT; 3158 AA.
AC  O52819;
DT  01-JUN-1998 (TrEMBLrel. 06, Created)
DT  01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT  01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE  PCZA363.3.
OS  Amycolatopsis orientalis.
OC  Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC  Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.
RN  [1]
RP  SEQUENCE FROM N.A.
RA  VAN WAGENINGEN A., KIRKPATRICK P., WILLIAMS D., HARRIS B., KERSHAW J.,
RL  Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR  EMBL: A0223999; CAA11794.1; -
DR  PROSITE: PS00455; AMP_BINDING; 3.
DR  PFAM: PF00501; AMP-binding; 3.
DR  PFAM: PF00568; DUF4; 3.
DR  PFAM: PF00550; pp-binding; 3.
DR  PRINTS: PR00154; AMPBINDING.
SQ  SEQUENCE 3158 AA; 336315 MW; E162473A CRC32;

Query Match      90.9%; Score 50; DB 2; Length 3158;
Best Local Similarity 62.5%; Pred. No. 4.52e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db  1055 GELSYGEL 1062
QY  139 SDSLSDGL 146
      :::::|

RESULT  4
ID  O30017      PRELIMINARY;      PRT; 281 AA.
AC  O30017;
DT  01-JAN-1998 (TrEMBLrel. 05, Created)
DT  01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT  01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE  BRANCHED-CHAIN AMINO ACID ABC TRANSPORTER, ATP-BINDING PROTEIN

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX  MEDLINE; 98049343.
RA  KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA  KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA  RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRIDES N.C.,
RA  FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA  KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA  PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA  OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA  COTTON M.D., SPRIGGS T., ARTIACH P., KATNE B.P., SYKES S.M.,
RA  SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA  MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESSE C.R.,
RA  VENTER J.C.;
RT  "The complete genome sequence of the hyperthermophilic, sulphate-
RT  reducing archaeon Archaeoglobus fulgidus.";
RL  Nature 390:364-370(1997).
DR  EMBL: AE001090; AAB91010.1; -
DR  TIGR: AF0222; -
DR  PFAM: PF00005; ABC_tran; 1.
KW  Hypothetical protein; ATP-binding.
SQ  SEQUENCE 281 AA; 31093 MW; ED7BF29D CRC32;

Query Match      89.1%; Score 49; DB 1; Length 281;
Best Local Similarity 62.5%; Pred. No. 7.40e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db  175 ANLSYSEL 182
QY  139 SDSLSDGL 146
      :::::|

RESULT  5
ID  O58324      PRELIMINARY;      PRT; 284 AA.
AC  O58324;
DT  01-AUG-1998 (TrEMBLrel. 07, Created)
DT  01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT  01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE  284AA LONG HYPOTHETICAL PROTEIN.
GN  PH0593.
OS  Pyrococcus horikoshii.
OC  Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=OT3;
RX  MEDLINE; 98344137.
RA  KAWARABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA  YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOIYAMA A., NAGAI Y.,
RA  SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA  FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA  AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA  KIKUCHI H.;
RT  "Complete sequence and gene organization of the genome of a hyper-
RT  thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL  DNA Res 5:55-76(1998).
DR  EMBL: AF000002; BAA29682.1; -
DR  PFAM: PF00892; DUF6; 1.
SQ  SEQUENCE 284 AA; 315631 MW; 6B5841B9 CRC32;

Query Match      87.3%; Score 48; DB 1; Length 284;
Best Local Similarity 85.7%; Pred. No. 1.20e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db  141 NLSYDGL 147
QY  140 SDSLSDGL 146
      :::::|

```

Thu May 11 06:50:02 2000

```

RESULT 6
ID O94421 PRELIMINARY; PRT; 430 AA.
AC O94421
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PUTATIVE TRANSCRIPTION REGULATORY PROTEIN-SNF2 FAMILY (FRAGMENT).
GN SPC1620.14C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-.
RA LYNE M., RAJANDREAM M.A., BARRELL B.G., HILBERT H., DUESTERHOEFF A.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL034491; CAA22498.1; -.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
FT NON_TER 430
SQ SEQUENCE 430 AA; 49624 MW; 2C613B568 CRC32;

Query Match 87.3%; Score 48; DB 3; Length 430;
Best Local Similarity 71.4%; Pred. No. 1.20e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 117 DLNYGDL 123
QY 140 DLSYGD 146
||:||||
||:||||

RESULT 7
ID P97260 PRELIMINARY; PRT; 1276 AA.
AC P97260;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE SREBP CLEAVAGE ACTIVATING PROTEIN.
GN SCAP.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RX HUA X., NOHTURFT A., GOLDSTEIN J.L., BROWN M.S.;
RA "sterol resistance in CHO cells traced to point mutation in SREBP cleavage-activating protein.";
RT Cell 87.415-426(1996).
RL EMBL; U67060; AAB19103.1; -.
DR PFAM; PF00400; WD40; 3.
DR PRINTS; PR00320; GPROTEINRPT.
SQ SEQUENCE 1276 AA; 139512 MW; FBD05617 CRC32;

Query Match 87.3%; Score 48; DB 11; Length 1276;
Best Local Similarity 85.7%; Pred. No. 1.20e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1223 DLNYGDL 1229
QY 140 DLSYGD 146
||:||||
||:||||

RESULT 8
ID Q12770 PRELIMINARY; PRT; 1277 AA.
AC Q12770;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE KIAA0199 PROTEIN (FRAGMENT).
GN KIAA0199.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96281124.
RA NAGASE T., SEKI N., ISHIKAWA K., TANAKA A., NOMURA N.;
RT "Prediction of the coding sequences of unidentified human genes. V. The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 3:17-24(1996).
DR EMBL; D83782; BAA12111.1; -.
DR PFAM; PF00400; WD40; 3.
FT NON_TER 1
SQ SEQUENCE 1277 AA; 139495 MW; 10EB5099 CRC32;

Query Match 87.3%; Score 48; DB 4; Length 1277;
Best Local Similarity 85.7%; Pred. No. 1.20e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1224 DLNYGDL 1230
QY 140 DLSYGD 146
||:||||
||:||||

RESULT 9
ID O26161 PRELIMINARY; PRT; 377 AA.
AC O26161;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
DE HYPOTHETICAL 42.2 KD PROTEIN.
GN MTH55.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Eumycarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RX MEDLINE; 98037514.
RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
RA HARRISON D., HOANG L., KENGLE P., LUMM W., POTIER B., QIU D.,
RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,
RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,
RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000797; AAB84562.1; -.
KW Hypothetical protein.
SQ SEQUENCE 377 AA; 42155 MW; 9A62CEC1 CRC32;

Query Match 85.5%; Score 47; DB 1; Length 377;
Best Local Similarity 62.5%; Pred. No. 1.94e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 179 GNLYGDL 186
QY 139 DLSYGD 146
||:||||
||:||||

RESULT 10
ID Q66080 PRELIMINARY; PRT; 194 AA.
AC Q66080;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBlrel. 01, Last annotation update)
DE TEGUMENT PROTEIN (FRAGMENT).
OS Canine herpesvirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae.

```

```

RN RP SEQUENCE FROM N.A.
RC STRAIN-MILOU;
RX MEDLINE: 96099411.
RA REMOND M., SHELDRIK P., LEBRETON F., NARDEUX P., FOULON T.;
RT "Directed integration of viral DNA mediated by fusion proteins
RT consisting of human immunodeficiency virus type 1 integrase and
RT Escherichia coli LexA protein."
RL J. Virol. 77:37-48(1996).
DR EMBL; X90449; CAA62075.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 194 AA; 22122 MW; 180D925C CRC32;

Query Match 83.6%; Score 46; DB 14; Length 194;
Best Local Similarity 75.0%; Pred. No. 3.11e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 131 TDSLXGTL 138
QY 139 SDSLXGDL 146
:|||||

RESULT 11
ID P74942 PRELIMINARY; PRT; 508 AA.
AC P74942;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE CLPB-HOMOLOGUE (FRAGMENT).
GN CLPB.
OS Thermus aquaticus (subsp. thermophilus).
OC Bacteria; Thermus/Deinococcus group; Thermus group; Thermus.
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN-HB8;
RX MEDLINE: 99194855.
RA KLOSTERMEIER D., SEIDEL R., REINSTEIN J.;
RT "The functional cycle and regulation of the Thermus thermophilus DnaK
RT chaperone system."
RL J. Mol. Biol. 287:511-525(1999).
DR EMBL; Y07826; CAA69163.2; -.
FT NON_TER 508
FT NON_TER 508
SQ SEQUENCE 508 AA; 57053 MW; 808F22AB CRC32;

Query Match 83.6%; Score 46; DB 2; Length 508;
Best Local Similarity 50.0%; Pred. No. 3.11e+01;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 490 AELRYGEL 497
QY 139 SDSLXGDL 146
:|||||

RESULT 12
ID Q65028 PRELIMINARY; PRT; 737 AA.
AC Q65028;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Arabis mosaic virus (AMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC Nepovirus.
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN-LILAC;
RX MEDLINE: 91341466.
RA BERTIOLI D.J., HARRIS R.D., EDWARDS M.L., COOPER J.I., HAWES W.S.;
RT "Transgenic plants and insect cells expressing the coat protein of
RT arabis mosaic virus produce empty virus-like particles."
RL J. Gen. Virol. 72:1801-1809(1991).
DR EMBL; D10086; BAA00982.1; -.

```

```

KW Polyprotein; Coat protein.
FT NON_TER 1
FT CHAIN 233 737 COAT PROTEIN.
SQ SEQUENCE 737 AA; 80881 MW; C1490F92 CRC32;

Query Match 83.6%; Score 46; DB 14; Length 737;
Best Local Similarity 71.4%; Pred. No. 3.11e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 339 DLDYGE 345
QY 140 DLSYGD 146
:|||||

RESULT 13
ID Q65029 PRELIMINARY; PRT; 1084 AA.
AC Q65029;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE POLYPROTEIN P2; 119KDA.
GN RNA2-L.
OS Arabis mosaic virus (AMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC Nepovirus.
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN-S;
RX MEDLINE: 97201550.
RA LOUDES A.M., RITZENTHALER C., PINCK M., SERGHINI M.A., PINCK L.;
RT "The 119 kDa and 124 kDa polyproteins of arabis mosaic nepovirus
RT (isolate S) are encoded by two distinct RNA2 species."
RL J. Gen. Virol. 76:899-906(1995).
DR EMBL; X81815; CAA57413.1; -.
KW Polyprotein; Coat protein.
FT CHAIN 1 233 N-TERMINAL 26KDA PROTEIN.
FT CHAIN 234 579 CENTRAL PUTATIVE MOVEMENT 38KDA PROTEIN.
FT CHAIN 580 1084 C-TERMINAL COAT 55.8KDA PROTEIN.
SQ SEQUENCE 1084 AA; 118997 MW; 496EC57E CRC32;

Query Match 83.6%; Score 46; DB 14; Length 1084;
Best Local Similarity 71.4%; Pred. No. 3.11e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 686 DLDYGE 692
QY 140 DLSYGD 146
:|||||

RESULT 14
ID Q9Y8M8 PRELIMINARY; PRT; 78 AA.
AC Q9Y8M8;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE 78AA LONG HYPOTHETICAL ACETYL-COA SYNTHETASE.
GN APES074.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN-K1;
RX MEDLINE: 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI H.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
DR EMBL; AF000064; BAA81622.1; -.

```

US-09-376-430-2-16.rspt

Thu May 11 06:50:02 2000

SQ SEQUENCE 78 AA; 8904 MW; A33A3544 CRC32;
 Query Match 81.8%; Score 45; DB 1; Length 78;
 Best Local Similarity 100.0%; Pred. No. 4.95e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 38 LSYGDL 43

|||||

QY 141 LSYGDL 146

RESULT 15
 ID O22368 PRELIMINARY; PRT; 155 AA.
 AC O22368;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE NACL-INDUCIBLE CA2+-BINDING PROTEIN.
 GN ACP1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA JANG H.J., PIH K.T., KANG S.G., LIM J.H., JIN J.B., PIAO H.L.,
 RA HWANG I.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF009228; AAC27657.1;
 DR HSSP; P02593; ITRC.
 DR PFAM; PF00036; eifhand; 3.
 SQ SEQUENCE 155 AA; 16760 MW; 82E90FB3 CRC32;

Query Match 81.8%; Score 45; DB 10; Length 155;
 Best Local Similarity 100.0%; Pred. No. 4.95e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 113 LSYGDL 118

|||||

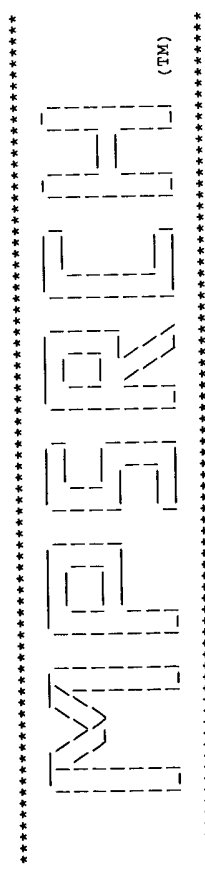
QY 141 LSYGDL 146

Search completed: Wed May 10 13:14:43 2000
 Job time : 240 secs.

24	55	37.9	332	1	R95449	Phage abortive infecti	1.81e+02
25	55	37.9	417	1	R67000	Staphylococcus epiderm	1.81e+02
26	55	37.9	536	1	R51195	Porcine zpa pellucida	1.81e+02
27	55	37.9	536	1	W81805	Porcine ZPB protein.	1.81e+02
28	55	37.9	541	1	W89547	Bacillus JPI70 proteas	1.81e+02
29	55	37.9	785	1	W98803	H. pylori GHPO 1302 pr	1.81e+02
30	55	37.9	38	1	R58457	TSAR binding domain en	2.26e+02
31	54	37.2	157	1	W88000	Protein encoded by exo	2.26e+02
32	54	37.2	229	1	W32621	Bacillus smithii nitri	2.26e+02
33	54	37.2	323	1	R35084	(Pro)leukocidin F.	2.26e+02
34	54	37.2	323	1	R35882	CH2 to CH4 of human Ig	2.26e+02
35	54	37.2	325	1	W88003	GA4 homologue GA4H1 fu	2.26e+02
36	54	37.2	405	1	R80277	E.coli L-carnitine den	2.26e+02
37	54	37.2	601	1	R88405	Human DNA-topoisomeras	2.26e+02
38	54	37.2	601	1	W41362	Human DNA topoisomeras	2.26e+02
39	54	37.2	1128	1	R49994	Mouse carboxypeptidase	2.26e+02
40	54	37.2	1128	1	W36817	Mouse E2A-binding prot	2.26e+02
41	54	37.2	1142	1	R70831	Human JAK1 kinase.	2.26e+02
42	54	37.2	1142	1	R25140	JAK1.	2.26e+02
43	54	37.2	3960	1	W56642	Protein kinase catalyt	2.26e+02
44	53	36.6	415	1	Y07061	Renal cancer associate	2.81e+02
45	53	36.6	934	1	W34451	Enterohaemorrhagic E.	2.81e+02

ALIGNMENTS

RESULT	1						
ID	W46873	standard; Protein; 554 AA.					
AC	W46873	11-JUN-1998 (first entry)					
DE	Bacillus thuringiensis toxin designated 573HD(a).						
KW	Toxin; lepidopteran pest; control; Agrotis ipsilon; black cutworm;						
KW	Heliothis virescens; Helicoverpa zea.						
OS	Bacillus thuringiensis.						
PH	Key	Location/Qualifiers					
FT	Misc_difference 26	/label= unknown					
FT	Misc_difference 27	/label= unknown					
FT	Misc_difference 29	/label= unknown					
FT	Misc_difference 33	/note= "encoded by MGG"					
FT	Misc_difference 37	/note= "encoded by CTS"					
FT	Misc_difference 41	/label= unknown					
FT	Misc_difference 46	/label= unknown					
FT	Misc_difference 58	/note= "encoded by TTR"					
FT	Misc_difference 68	/label= unknown					
FT	Misc_difference 80	/label= unknown					
FT	Misc_difference 93	/label= unknown					
FT	Misc_difference 99	/label= unknown					
FT	Misc_difference 105	/label= unknown					
FT	Misc_difference 119	/label= unknown					
FT	Misc_difference 141	/label= unknown					
FT	Misc_difference 279	/label= unknown					
FT	W09800546-A2.	/label= unknown					
PN	08-JAN-1998.						
PD	01-JUL-1997; U11658.						
PR	01-JUL-1996; US-674002.						
PA	(MYCO) MYCOGEN CORP.						
PI	Narva KE, Schnepf HE, Stockhoff BA, Walz M, Wicker C;						



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPPerch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 13:23:44 2000; MasPar time 3.19 Seconds
141.102 Million cell updates/sec
Tabular output not generated.

Title: >US-09-376-430-2
Description: (151-169) from US09376430A.pep (17 of 25)
Perfect Score: 145
Sequence: 1 QYRSPFDTEWQSKQENTCN 19
Scoring table: PAM 150
Gap 11
Searched: 188963 seqs, 23686106 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: a-geneseq35
l:geneseqp
Statistics: Mean 20.742; Variance 65.295; scale 0.318

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	61	42.1	554	1	W46873	Bacillus thuringiensis
2	61	42.1	759	1	W46864	Bacillus thuringiensis
3	59	40.7	762	1	Y00939	Desulfurococcus M11 TL
4	59	40.7	1028	1	W29667	Homo sapiens DL185_1 c
5	58	40.0	94	1	R47117	Complete sequence of h
6	58	40.0	389	1	W27138	Mouse alpha-fucosyltra
7	58	40.0	533	1	W83317	Mouse Lrp5 protein fro
8	58	40.0	1451	1	W83310	Lrp5 protein from isof
9	58	40.0	1591	1	W83308	Mature Lrp5 protein.
10	58	40.0	1614	1	W83312	Mouse Lrp5 protein.
11	58	40.0	1615	1	W83309	Lrp5 protein from the
12	58	40.0	1639	1	W83311	Lrp5 isoform 3 protein
13	57	39.3	527	1	W47207	Homo sapiens tubulin-f
14	57	39.3	691	1	R38735	Beta-galactosidase.
15	56	38.6	14	1	W01827	CD4 epitope.
16	56	38.6	143	1	R49565	Fusion protein JB122.
17	56	38.6	256	1	W98653	H. pylori GHPO 439 pro
18	56	38.6	503	1	R14768	Metastasis-specific va
19	56	38.6	636	1	W89548	Bacillus sp. alkaline
20	56	38.6	1433	1	R80474	Sequence of protein of
21	55	37.9	61	1	R32224	Partial fibroblast gro
22	55	37.9	207	1	W62623	Homo sapiens SOCS11 pr
23	55	37.9	325	1	R77862	Clavaminase-synthase.

DR WPI: 98-086971/08.
 DR N-PSDB: V16532.
 PT New isolated *Bacillus thuringiensis* isolate(s) - used to obtain
 PI genes encoding toxins which are active against lepidopteran pests
 PS such as the Black cutworm
 PS Claim 42; Pages 167-169; 183pp; English.
 CC The present sequence represents a *Bacillus thuringiensis* toxin
 CC which is active against lepidopteran pests. The toxin isolates
 CC can be used for the control of lepidopteran pests such as *Agrotis*
 CC *ipsilon* (black cutworm), *Heliothis virescens* and *Helicoverpa zea*.
 CC PCR primers and probes can be derived from the polynucleotide
 CC encoding the toxin and used for the amplification and detection of
 CC other toxin-encoding sequences.
 SQ Sequence 554 AA;

Query Match 42.1%; Score 61; DB 1; Length 554;
 Best Local Similarity 47.1%; Pred. No. 4.61e+01;
 Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

DB 370 RRNFIIYWRVKTENVN 386
 QY 153 RSPFDTEWQSKQENTCN 169

RESULT 2

ID W46864 standard; Protein; 759 AA.
 AC W46864;
 DT 11-JUN-1998 (first entry)
 DE *Bacillus thuringiensis* toxin designated 86W1(a).
 KW Toxin, lepidopteran pest; control; *Agrotis ipsilon*; black cutworm;
 OS *Heliothis virescens*; *Helicoverpa zea*.
 FH *Bacillus thuringiensis*.
 FT Key Location/Qualifiers
 FT Misc_difference 193
 FT /label= unknown
 FT /note= "encoded by AKT"
 FT Misc_difference 257
 FT /label= unknown
 FT /note= "encoded by AST"
 FT W09800546-A2.
 PD 08-JAN-1998.
 PF 01-JUL-1997; U11658.
 PR 01-JUL-1996; US-674002.
 PA (MYCO) MYCOGEN CORP.
 PI Narva KE, Schnepf HE, Stockhoff BA, Walz M, Wicker C;
 DR WPI: 98-086971/08.
 DR N-PSDB: V16523.
 PT New isolated *Bacillus thuringiensis* isolate(s) - used to obtain
 PT genes encoding toxins which are active against lepidopteran pests
 PT such as the Black cutworm
 PS Claim 42; Pages 126-129; 183pp; English.
 CC The present sequence represents a *Bacillus thuringiensis* toxin
 CC which is active against lepidopteran pests. The toxin isolates
 CC can be used for the control of lepidopteran pests such as *Agrotis*
 CC *ipsilon* (black cutworm), *Heliothis virescens* and *Helicoverpa zea*.
 CC PCR primers and probes can be derived from the polynucleotide
 CC encoding the toxin and used for the amplification and detection of
 CC other toxin-encoding sequences.
 CC note: nucleotide sequence contains numerous stop codons which do not
 CC alter the reading frame.
 SQ Sequence 759 AA;

Query Match 42.1%; Score 61; DB 1; Length 759;
 Best Local Similarity 47.1%; Pred. No. 4.61e+01;
 Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

DB 575 RRNFIIYWRVKTENVN 591
 QY 153 RSPFDTEWQSKQENTCN 169

RESULT 3

ID Y00939 standard; Protein; 762 AA.

AC Y00939;
 DT 02-JUN-1999 (first entry)
 DE *Desulfurococcus* M11 TL DNA polymerase 29pY1 protein sequence.
 KW DNA polymerase; thermophilic bacteria; DNA synthesis.
 OS *Desulfurococcus* sp.
 PN WO9907837-A1.
 PD 18-FEB-1999.
 PF 06-AUG-1998; U17152.
 PR 06-AUG-1997; US-907166.
 PA (DIVE-) DIVERSA INC.
 PI Callen W, Mathur EJ;
 DR WPI: 99-180490/15.
 DR N-PSDB: X27286.
 PT DNA polymerases from extremely thermophilic bacteria - useful for
 PT DNA synthesis
 PS Claim 1; Fig 5; 72pp; English.
 CC This sequence is a DNA polymerase of the invention, that was
 CC isolated from a thermophilic bacteria. The polymerases are used in DNA
 CC synthesis and as immunogens to raise antibodies (useful for affinity
 CC purification and to screen for related enzymes). Fragments of the DNA
 CC encoding the polymerases are used as probes to isolate related or
 CC full-length sequences and to produce the recombinant polymerases. The
 CC polymerases catalyze DNA synthesis by the addition of deoxynucleotides to
 CC the 3' end of a polynucleotide chain, using a complementary
 CC polynucleotide strand as a template. The polymerases have optimum
 CC activity at over 60 deg. C and can renature and regain activity after
 CC exposure to temperatures above 70 deg. C.
 SQ Sequence 762 AA;

Query Match 40.7%; Score 59; DB 1; Length 762;
 Best Local Similarity 54.5%; Pred. No. 7.31e+01;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 21 RSKFNATAWEAK 31
 QY 153 RSPFDTEWQSK 163

RESULT 4

ID W29667 standard; Protein; 1028 AA.
 AC W29667;
 DT 27-OCT-1998 (first entry)
 DE Homo sapiens DL185_1 clone secreted protein.
 KW secreted protein; DL185_1.
 OS Homo sapiens.
 PN W09830695-A2.
 PD 16-JUL-1998.
 PF 09-JAN-1998; U00543.
 PR 08-JAN-1998; US-004684.
 PR 09-JAN-1997; US-780814.
 PA (GEM) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, Lavallic ER, McCoy JM, Merberg D,
 PI Racie LA, Spaulding V, Treacy M;
 DR WPI: 98-413686/35.
 DR N-PSDB: V40529.
 PT New isolated nucleic acids and secreted proteins - obtained from
 PT human adult ovary, human foetal kidney, human foetal brain and human
 PT adult brain cDNA libraries
 PS Claim 36; Page 91-94; 113pp; English.
 CC The sequence is that of a novel, isolated secreted protein.
 SQ Sequence 1028 AA;

Query Match 40.7%; Score 59; DB 1; Length 1028;
 Best Local Similarity 50.0%; Pred. No. 7.31e+01;
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

DB 634 QTRTPFSVGWQA 645
 QY 151 QYRSPFDTEWQSK 162

RESULT 5

ID R47117 standard; peptide; 94 AA.

RESULT	6	
ID	W27138	standard; Protein; 389 AA.
AC	02-MAR-1998	(first entry)
DE	Mouse alpha-fucosyltransferase Fuc-TVII (44.5 kDa).	
DE	Mouse alpha-fucosyltransferase; Fuc-TVII gene; mouse;	
KW	Alpha-fucosyltransferase; Fuc-TVII gene; mouse;	
KW	sialyl Lewis x determinant; oligosaccharide; antiinflammatory;	
KW	inhibitor.	
MS	Mus musculus.	
PN	M09732889-A1.	
PD	12-SEP-1997.	
PF	07-MAR-1997;	U03184.
PR	08-MAR-1996;	US-613098.
PA	(UNWI) UNIV MICHIGAN.	
PI	Gersten KM, Lowe JB, Natsuka S;	
DR	WPI; 97-470335/43.	
DR	N-PSDB; T85219.	
PT	Novel murine alpha-(1-3)-fucosyl-transferase(s) - useful for e.g.	
PT	screening for inhibitors used as antiinflammatory agents	
PS	Claim 1; Fig 2; 91pp; English.	
CC	This protein comprises a 44.5 kDa mouse alpha-fucosyltransferase	
CC	encoded by nucleotides 996-1149 and 2067-3079 of a Fuc-TVII gene	
CC	sequence (see T85219) isolated from a 3T3 cell genomic DNA library.	
CC	The enzyme has one of two amino acid sequences (see also W26671)	
CC	resulting from structurally distinct transcripts derived from	
CC	alternative splicing and alternative transcription initiation events.	
CC	Fuc-TVII proteins and nucleic acids can be used to construct animal	
CC	cell lines able to post-translationally modify oligosaccharides on	
CC	cell surface, intracellular or secreted proteins and lipids, to	
CC	isolate reagents for efficient enzymatic production of	
CC	oligosaccharides, to generate antibodies to glycosyltransferases	
CC	useful as diagnostic reagents, to screen for fucosyltransferase	
CC	inhibitors and inactivators, especially those that act as	
CC	antiinflammatory agents, for genotyping individuals at the	
CC	fucosyltransferase locus and for in vitro synthesis of sialyl Lewis	
CC	x tetrasaccharide.	

SQL	Sequence	389 AA;
	Query Match	40.0%; Score 58; DB 1; Length 389;
	Best Local Similarity	60.0%; Pred. No. 9.18e+01;
	Matches	6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db	29 WESRKEATCN 38	
	.: : :	
QY	160 WQSQKQTCN 169	
	:	
RESULT	7	
ID	W83317 standard; Protein: 533 AA.	
AC	W83317;	
DT	10-FEB-1999 (first entry)	
DE	Mouse Lrp5 protein from a portion of the open reading frame.	
KW	LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;	
KW	endocytosis; insulin dependent diabetes mellitus; autoimmune disease;	
KW	glomerulonephritis; inflammation; viral infection; osteoporosis;	
KW	hypercholesterolemia; Alzheimer's disease; low density lipoprotein.	
OS	Mus sp.	
PX	WC09846743-A1.	
PD	22-OCT-1998.	
PF	15-APR-1998; G01102.	
PR	05-JUN-1997; US-048740.	
PR	15-APR-1997; US-043553.	
PA	(MERI) MERCK & CO INC.	
PA	(WELL) WELLCOME TRUST LTD.	
PI	Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW,	
PI	Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,	
PI	Phillips MS, Todd JA, Twells RCJ;	
DR	WPI; 98-594573/50.	
PT	New isolated LDL-receptor related protein - used to develop products	
PT	for treating, e.g. elevated triglyceride levels, diabetes,	
PT	autoimmune disorders, inflammation or Alzheimer's disease	
PS	Example 1: Fig 16d: 200pp: English.	
CC	The present invention describes LRP5 (low density lipoprotein (LDL)	
CC	receptor related protein, previously designated LRP-3). The present	
CC	sequence represents a the mouse Lrp5 protein.	
CC	Nucleic acid molecules (NAMS) encoding LRP5 can be used for determining	
CC	if an individual is susceptible to insulin dependent diabetes mellitus	
CC	(IDDM). The NAMS or proteins can be used for reducing triglyceride levels	
CC	in the serum of an individual. Therapies that affect LRP5 may also be	
CC	useful in the treatment of autoimmune diseases such as	
CC	glomerulonephritis, diseases and disorders involving disruption of	
CC	endocytosis and/or antigen presentation, cytokine clearance and/or	
CC	inflammation, viral infection, pathogenic bacterial toxin contamination,	
CC	elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,	
CC	osteoporosis, Alzheimer's disease and cardiovascular disease. Products	
CC	from the present invention can also be used for detection, diagnosis and	
CC	drug screening. AA;	
SQL	Sequence	533 AA;
	Query Match	40.0%; Score 58; DB 1; Length 533;
	Best Local Similarity	41.7%; Pred. No. 9.18e+01;
	Matches	5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
Db	252 TDWQRSIHACN 263	
	: ::: :	
QY	158 TEWSQKQTCN 169	
RESULT	8	
ID	W83310 standard; Protein: 1451 AA.	
AC	W83310;	
DT	10-FEB-1999 (first entry)	
DE	LRP5 protein from isoform 2 (also isoform 4,5,6).	
KW	LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;	
KW	endocytosis; insulin dependent diabetes mellitus; autoimmune disease;	
KW	glomerulonephritis; inflammation; viral infection; osteoporosis;	
KW	hypercholesterolemia; Alzheimer's disease; low density lipoprotein.	
OS	Homo sapiens.	
PX	WC09846743-A1.	

PD 22-OCT-1998.
 PF 15-APR-1998; G01102.
 PR 05-JUN-1997; US-048740.
 PR 15-APR-1997; US-043553.
 PA (MERI) MERCK & CO INC.
 PA (WELL) WELLCOME TRUST LTD.
 PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW,
 PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
 PI Phillips MS, Todd JA, Twells RCJ;
 DR WPI: 98-594573/50.
 PT New isolated LDL-receptor related protein - used to develop products
 PT for treating, e.g. elevated triglyceride levels, diabetes,
 PT autoimmune disorders, inflammation or Alzheimer's disease
 PS Claim 6; Fig 11c; 200pp; English.
 CC The present invention describes LRP5 (low density lipoprotein (LDL)
 CC receptor related protein, previously designated LRP-3). The present
 CC sequence represents the protein from LRP5 isoform 2 (also isoform 4,5,6).
 CC Nucleic acid molecules (NAMS) encoding LRP5 can be used for determining
 CC if an individual is susceptible to insulin dependent diabetes mellitus
 CC (IDDM). The NAMS or proteins can be used for reducing triglyceride levels
 CC in the serum of an individual. Therapies that affect LRP5 may also be
 CC useful in the treatment of autoimmune diseases such as
 CC glomerulonephritis, diseases and disorders involving disruption of
 CC endocytosis and/or antigen presentation, cytokine clearance and/or
 CC inflammation, viral infection, pathogenic bacterial toxin contamination,
 CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
 CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
 CC from the present invention can also be used for detection, diagnosis and
 CC drug screening.
 SQ Sequence 1451 AA;

Query Match 40.0%; Score 58; DB 1; Length 1451;
 Best Local Similarity 41.7%; Pred. No. 9.18e+01;
 Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 Db 89 TDQWTRSIHACN 100
 QY 158 TEWQSKQENTCN 169

RESULT 9
 ID W83308 standard; protein; 1591 AA.
 AC W83308.
 DT 10-FEB-1999 (first entry)
 DE Mature Lrp5 protein.
 KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
 KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
 KW glomerulonephritis; inflammation; viral infection; osteoporosis;
 KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein.
 OS Homo sapiens.
 PN W09846743-AL.
 PD 22-OCT-1998.
 PF 15-APR-1998; G01102.
 PR 05-JUN-1997; US-048740.
 PR 15-APR-1997; US-043553.
 PA (MERI) MERCK & CO INC.
 PA (WELL) WELLCOME TRUST LTD.
 PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW,
 PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
 PI Phillips MS, Todd JA, Twells RCJ;
 DR WPI: 98-594573/50.
 PT New isolated LDL-receptor related protein - used to develop products
 PT for treating, e.g. elevated triglyceride levels, diabetes,
 PT autoimmune disorders, inflammation or Alzheimer's disease
 PS Claim 1; Fig 5e; 200pp; English.
 CC The present sequence represents mature LRP5 protein (low density
 CC lipoprotein (LDL) receptor related protein, previously designated LRP-3).
 CC Nucleic acid molecules (NAMS) encoding LRP5 can be used for determining
 CC if an individual is susceptible to insulin dependent diabetes mellitus
 CC (IDDM). The NAMS or proteins can be used for reducing triglyceride levels
 CC in the serum of an individual. Therapies that affect LRP5 may also be
 CC useful in the treatment of autoimmune diseases such as
 CC glomerulonephritis, diseases and disorders involving disruption of

CC endocytosis and/or antigen presentation, cytokine clearance and/or
 CC inflammation, viral infection, pathogenic bacterial toxin contamination,
 CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
 CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
 CC from the present invention can also be used for detection, diagnosis and
 CC drug screening.
 SQ Sequence 1591 AA;

Query Match 40.0%; Score 58; DB 1; Length 1591;
 Best Local Similarity 41.7%; Pred. No. 9.18e+01;
 Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 Db 229 TDQWTRSIHACN 240
 QY 158 TEWQSKQENTCN 169

RESULT 10
 ID W83312 standard; protein; 1614 AA.
 AC W83312.
 DT 10-FEB-1999 (first entry)
 DE Mouse Lrp5 protein.
 KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
 KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
 KW glomerulonephritis; inflammation; viral infection; osteoporosis;
 KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein.
 OS Mus sp.
 PN W09846743-AL.
 PD 22-OCT-1998.
 PF 15-APR-1998; G01102.
 PR 05-JUN-1997; US-048740.
 PR 15-APR-1997; US-043553.
 PA (MERI) MERCK & CO INC.
 PA (WELL) WELLCOME TRUST LTD.
 PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW,
 PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
 PI Phillips MS, Todd JA, Twells RCJ;
 DR WPI: 98-594573/50.
 PT New isolated LDL-receptor related protein - used to develop products
 PT for treating, e.g. elevated triglyceride levels, diabetes,
 PT autoimmune disorders, inflammation or Alzheimer's disease.
 PS Claim 6; Fig 18c; 200pp; English.
 CC The present invention describes LRP5 (low density lipoprotein (LDL)
 CC receptor related protein, previously designated LRP-3). The present
 CC sequence represents the mouse Lrp5 protein.
 CC Nucleic acid molecules (NAMS) encoding LRP5 can be used for determining
 CC if an individual is susceptible to insulin dependent diabetes mellitus
 CC (IDDM). The NAMS or proteins can be used for reducing triglyceride levels
 CC in the serum of an individual. Therapies that affect LRP5 may also be
 CC useful in the treatment of autoimmune diseases such as
 CC glomerulonephritis, diseases and disorders involving disruption of
 CC endocytosis and/or antigen presentation, cytokine clearance and/or
 CC inflammation, viral infection, pathogenic bacterial toxin contamination,
 CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
 CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
 CC from the present invention can also be used for detection, diagnosis and
 CC drug screening.
 SQ Sequence 1614 AA;

Query Match 40.0%; Score 58; DB 1; Length 1614;
 Best Local Similarity 41.7%; Pred. No. 9.18e+01;
 Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 Db 252 TDQWTRSIHACN 263
 QY 158 TEWQSKQENTCN 169

RESULT 11
 ID W83309 standard; protein; 1615 AA.
 AC W83309.
 DT 10-FEB-1999 (first entry)
 DE LRP5 protein from the longest open reading frame.

CC LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
CC endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
CC glomerulonephritis; inflammation; viral infection; osteoporosis;
CC hypercholesterolemia; Alzheimer's disease; low density lipoprotein.
OS Homo sapiens.
PN WO9846743-A1.
PD 22-OCT-1998.
PF 15-APR-1998; G01102.
PR 05-JUN-1997; US-048740.
PR 15-APR-1997; US-043553.
PR 15-APR-1997; US-043553.
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW, Nakagawa Y,
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
PI Phillips MS, Todd JA, Twells RCJ;
DR WPI; 98-594573/50.
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
PS Claim 2; Fig 5c; 200pp; English.
CC The present sequence represents a LRP5 protein (low density lipoprotein
CC (LDL) receptor related protein, previously designated LRP-3).
CC Nucleic acid molecules (NAMS) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus
CC (IDDM). The NAMS or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
SQ Sequence 1639 AA;
Query Match 40.0%; Score 58; DB 1; Length 1639;
Best Local Similarity 41.7%; Pred. No. 9.18e+01;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
Db 277 TDWOTRSIHACN 288
QY 158 TEMQSKQENTCN 169
RESULT 13
ID W47207 standard; Protein; 527 AA.
AC W47207;
DT 21-JUL-1998 (first entry)
DE Homo sapiens tubulin-folding cofactor E.
KW Tubulin folding; cofactor; alpha-tubulin; beta-tubulin; unfolded;
KW folded; treatment; hyper-proliferative diseases; cancer; ss.
OS Homo sapiens.
PN WO9804587-A1.
PD 05-FEB-1998.
PF 25-JUL-1997; U14076.
PR 25-JUL-1996; US-023089.
PR (UUNY) UNIV NEW YORK STATE.
PI Cowan NJ;
PI WPI; 98-130618/12.
DR N-PSDB; V17087.
DT New isolated cofactor(s) for tubulin folding - are useful as targets
PT for identifying agents which interfere with folding in the treatment
PT of hyper-proliferative diseases such as cancer
PS Claim 4; Pages 52-54; 87pp; English.
CC The sequence is that encoding human tubulin-folding cofactor E.
CC It may be useful as a target for interfering with the
CC production of productively folded alpha- and beta-tubulins.
CC Since tubulin function is essential for cell division and
CC proliferation, agents which interfere with tubulin function
CC can serve as useful antiproliferative compounds. Such interfering
CC agents have potential utility as agents for the treatment of
CC hyperproliferative diseases such as cancer and the treatment
CC of gout.
SQ Sequence 527 AA;
Query Match 39.3%; Score 57; DB 1; Length 527;
Best Local Similarity 50.0%; Pred. No. 1.15e+02;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Db 384 DYKAFGNEW 393
QY 151 QYRSPFDIEW 160
RESULT 14
ID R38735 standard; Protein; 691 AA.
AC R38735;
DT 29-NOV-1993 (first entry)
DE Beta-galactosidase.
KW Beta-galactosidase; B. breve; YIT4010; hydrolase; E. coli; lactose;
KW transformation; recombinant plasmid; galacto-oligo sugar; pBR322.
OS Bifidobacterium breve.
PN J05146296-A.
PD 15-JUN-1993.

KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein.
OS Homo sapiens.
PN WO9846743-A1.
PD 22-OCT-1998.
PF 15-APR-1998; G01102.
PR 05-JUN-1997; US-048740.
PR 15-APR-1997; US-043553.
PR 15-APR-1997; US-043553.
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW, Nakagawa Y,
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
PI Phillips MS, Todd JA, Twells RCJ;
DR WPI; 98-594573/50.
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
PS Claim 2; Fig 5c; 200pp; English.
CC The present sequence represents a LRP5 protein (low density lipoprotein
CC (LDL) receptor related protein, previously designated LRP-3).
CC Nucleic acid molecules (NAMS) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus
CC (IDDM). The NAMS or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
SQ Sequence 1615 AA;
Query Match 40.0%; Score 58; DB 1; Length 1615;
Best Local Similarity 41.7%; Pred. No. 9.18e+01;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
Db 253 TDWOTRSIHACN 264
QY 158 TEMQSKQENTCN 169
RESULT 12
ID W83311 standard; Protein; 1639 AA.
AC W83311;
DT 10-FEB-1999 (first entry)
DE LRP5 isoform 3 protein.
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein.
OS Homo sapiens.
PN WO9846743-A1.
PD 22-OCT-1998.
PF 15-APR-1998; G01102.
PR 05-JUN-1997; US-048740.
PR 15-APR-1997; US-043553.
PR 15-APR-1997; US-043553.
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW, Nakagawa Y,
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
PI Phillips MS, Todd JA, Twells RCJ;
DR WPI; 98-594573/50.
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
PS Claim 6; Fig 12d; 200pp; English.
CC The present invention describes LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3). The present
CC sequence represents the protein from LRP5 isoform 3.

PF 27-NOV-1991; 335904.
PR 27-NOV-1991; JP-335904.
PA (HONS) YAKULT HONSHA KK.
DR WPI: 93-223533/28.
DR N-PSDB; 046333.
PT New DNA fragment - contg. beta-galactosidase
PS Claim 1; Page 7-10; 12pp; Japanese.
CC This sequence is encoded by the beta-galactosidase gene derived from
CC Bifidobacterium breve YIN4010. The DNA fragment was digested with
CC BamHI and ligated into BamHI digested pBR322. E. coli lacking beta-
CC galactosidase activity were transformed with the recombinant plasmid.
CC Transformsants having beta-galactosidase activity were cultured and
CC the beta-galactosidase gene was isolated. The gene may be used in
CC the production of large amounts of B. brevis beta-galactosidase in
CC conjunction with E. coli beta-galactosidase hydrolases, lactose and
CC galacto-oligo sugars.

SQ Sequence 691 AA;

Query Match 39.3%; Score 57; DB 1; Length 691;
Best Local Similarity 25.0%; Pred. No. 1.15e+02;
Matches 5; Conservative 6; Mismatches 8; Indels 1; Gaps 1;

Db 145 HYKDNPVVVSWHWYSNEYGCH 164
 :|: |:|: |:: ||:
QY 151 QYR-SPPDTEWQSKQENTCN 169

RESULT 15
ID W01827 standard; peptide; 14 AA.
AC W01827;
DT 20-JUL-1997 (first entry)
DE CD4 epitope.
KW T-cell; receptor; CD44; epitope; fusion protein; cancer;
KW Immunoglobulin.
OS Rattus rattus.
PN DE19540515-C1.
PD 06-FEB-1997.
PF 31-OCT-1995; 040515.
PR 31-OCT-1995; DE-040515.
PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.
PA (GESL) FORSCHUNGSZENTRUM KARLSRUHE GMBH.
PI Hekele A., Herrlich P., Ponta H.;
PR WPI: 97-101136/10.
PT T cells expressing fusion protein specific for variant CD44 gene -
PT useful for cancer therapy, esp. pancreatic carcinoma
PS Claim 6; Page 9; 16pp; English.
CC The fusion protein (T62573) comprises a first portion having
CC specific affinity for an amino acid sequence encoded by a variant
CC exon of the CD44 gene [pref. W01827 [rat] or W01828 [human]], and a
CC second portion comprising at least part of the amino acid sequence of
CC a subunit of the T-cell receptor complex or of an immunoglobulin
CC receptor. Products contg. the fusion protein (or DNA encoding it)
CC are useful for treating cancer, esp. mammary, colonic, gastric or
CC pancreatic carcinoma, and metastatic disease.

SQ Sequence 14 AA;

Query Match 38.6%; Score 56; DB 1; Length 14;
Best Local Similarity 50.0%; Pred. No. 1.44e+02;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Search completed: Wed May 10 13:23:51 2000
Job time : 7 secs.

WIPREH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 13:24:09 2000; MasPar time 44.88 Seconds

Tabular output not generated. 5.485 Million cell updates/sec

Title: >US-09-376-430-2
Description: (151-169) from US09376430A.pep (17 of 25)
Perfect Score: 145
Sequence: 1 QYRSPFDTEWQSKQENTCN 19

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 19.587; Variance 64.718; scale 0.303

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	DB ID	Description	Pred. No.
1	61	42.1	114	4	5514582-13 Patent No. 5514582.	2.74e+01
2	59	40.7	762	2	US-08-907- Sequence 10, Applicati	4.30e+01
3	58	40.0	94	1	US-07-899- Sequence 1, Applicatio	5.38e+01
4	56	38.6	14	2	US-08-750- Sequence 3, Applicatio	8.39e+01
5	56	38.6	42	2	US-08-750- Sequence 5, Applicatio	8.39e+01
6	56	38.6	334	2	US-08-359- Sequence 4, Applicatio	8.39e+01
7	56	38.6	355	1	US-07-946- Sequence 5, Applicatio	8.39e+01
8	56	38.6	355	1	US-08-478- Sequence 5, Applicatio	8.39e+01
9	56	38.6	355	1	US-08-483- Sequence 5, Applicatio	8.39e+01
10	56	38.6	503	1	US-07-946- Sequence 2, Applicatio	8.39e+01
11	56	38.6	503	1	US-08-483- Sequence 2, Applicatio	8.39e+01
12	56	38.6	503	2	US-08-478- Sequence 2, Applicatio	8.39e+01
13	55	37.9	536	2	US-08-484- Sequence 4, Applicatio	1.05e+02
14	55	37.9	536	2	US-08-484- Sequence 4, Applicatio	1.05e+02
15	55	37.9	536	2	US-08-480- Sequence 4, Applicatio	1.05e+02
16	55	37.9	536	2	US-08-484- Sequence 4, Applicatio	1.05e+02
17	55	37.9	641	2	US-08-873- Sequence 42, Applicatio	1.05e+02
18	55	37.9	3169	2	US-08-477- Sequence 6, Applicatio	1.05e+02
19	54	37.2	38	1	US-08-471- Sequence 85, Applicati	1.30e+02
20	54	37.2	38	1	US-08-189- Sequence 85, Applicati	1.30e+02
21	54	37.2	38	2	US-08-471- Sequence 85, Applicati	1.30e+02
22	54	37.2	38	2	US-08-471- Sequence 85, Applicati	1.30e+02
23	54	37.2	38	2	US-08-471- Sequence 85, Applicati	1.30e+02

24	54	37.2	38	1	US-08-176- Sequence 85, Applicati	1.30e+02
25	54	37.2	169	2	US-08-946- Sequence 8, Applicatio	1.30e+02
26	54	37.2	501	2	US-08-660- Sequence 13, Applicati	1.30e+02
27	54	37.2	525	2	US-08-786- Sequence 2, Applicatio	1.30e+02
28	54	37.2	525	1	US-08-356- Sequence 2, Applicatio	1.30e+02
29	54	37.2	601	1	US-08-458- Sequence 2, Applicatio	1.30e+02
30	54	37.2	601	2	US-09-033- Sequence 2, Applicatio	1.30e+02
31	54	37.2	1128	1	US-08-111- Sequence 2, Applicatio	1.30e+02
32	54	37.2	1132	1	US-08-446- Sequence 18, Applicati	1.30e+02
33	54	37.2	1132	1	US-08-805- Sequence 18, Applicati	1.30e+02
34	54	37.2	1132	1	US-08-446- Sequence 18, Applicati	1.30e+02
35	54	37.2	1132	2	US-08-064- Sequence 18, Applicati	1.30e+02
36	54	37.2	1132	2	US-09-066- Sequence 18, Applicati	1.30e+02
37	54	37.2	1142	1	US-08-097- Sequence 11, Applicati	1.30e+02
38	54	37.2	1153	1	US-08-097- Sequence 14, Applicati	1.30e+02
39	54	37.2	1154	1	US-08-357- Sequence 7, Applicatio	1.30e+02
40	54	37.2	1154	3	PCT-US95-1 Sequence 7, Applicatio	1.30e+02
41	54	37.2	1154	1	US-08-446- Sequence 24, Applicati	1.30e+02
42	54	37.2	1154	2	US-09-003- Sequence 7, Applicatio	1.30e+02
43	54	37.2	3959	2	US-08-970- Sequence 30, Applicati	1.30e+02
44	53	36.6	740	1	US-08-781- Sequence 12, Applicati	1.62e+02
45	53	36.6	750	2	US-08-820- Sequence 4, Applicatio	1.62e+02

ALIGNMENTS

RESULT 1
ID 5514582-13 STANDARD; PRT; 124 AA.
XX
AC xxxxxx
DT 01-JAN-1900
DE Patent No. 5514582.
XX
CC Patent No. 5514582
CC APPLICANT: CAPON, DANIEL J.; LASKY, LAURENCE A.
CC TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID
CC IMMUNOGLOBULINS
CC NUMBER OF SEQUENCES: 43
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/185,670
CC FILING DATE: 21-JAN-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 986,931
CC FILING DATE: 08-DEC-1992
CC APPLICATION NUMBER: 808,122
CC FILING DATE: 16-DEC-1991
CC APPLICATION NUMBER: 440,625
CC FILING DATE: 22-NOV-1989
CC APPLICATION NUMBER: 315,015
CC FILING DATE: 23-FEB-1989
CC SEQ ID NO:13
CC LENGTH: 114
SQ SEQUENCE 124 AA; 14459 MW; 79120 CN;

Query Match 42.1%; Score 61; DB 4; Length 114;
Best Local Similarity 38.9%; Pred. No. 2.74e+01;
Matches 7; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

Db 65 EYRSNF-KNWAFTQDNC 81
QY 151 QYRSPFDTEWQSKQENTC 168

RESULT 2
ID US-08-907-166-10 STANDARD; PRT; 762 AA.
XX
AC xxxxxx

Sequence 10, Application US/08907166

XX Sequence 10, Application US/08907166
 CC Patent No. 5948666
 CC GENERAL INFORMATION:
 CC APPLICANT: Callen, Walter
 CC APPLICANT: Mather, Eric
 CC TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
 CC FILE REFERENCE: 09010/027001
 CC CURRENT APPLICATION NUMBER: US/08/907,166
 CC CURRENT FILING DATE: 1997-08-06
 CC NUMBER OF SEQ ID NOS: 12
 CC SOFTWARE: FastSeq for Windows Version 3.0
 CC SEQ ID NO 10
 CC LENGTH: 762
 CC TYPE: PRT
 CC ORGANISM: Desulfurococcus sp.
 CC FEATURE:
 CC NAME/KEY: variation
 CC LOCATION: (601)..(601)
 CC OTHER INFORMATION: Xaa at position 601 is alanine or proline
 CC SEQUENCE 762 AA; 87246 MW; 3129314 CN;
 CC
 CC Query Match 40.7%; Score 59; DB 2; Length 762;
 CC Best Local Similarity 54.5%; Pred. No. 4.30e+01;
 CC Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 CC
 Db 21 RSKFNTAEAK 31
 YY || || || || ||
 QY 153 RSPFDTEWQSK 163
 CC
 CC RESULT 3
 CC ID US-07-899-535A-1 STANDARD; PRT; 94 AA.
 CC XX
 CC AC xxxxxx
 CC DT
 CC DT
 CC DT
 CC DE Sequence 1, Application US/07899535A
 CC Sequence 1, Application US/07899535A
 CC Patent No. 5428011
 CC GENERAL INFORMATION:
 CC APPLICANT: Sheth, Anil R.
 CC APPLICANT: Garde, Seema
 CC APPLICANT: Panchal, Chandra J.
 CC TITLE OF INVENTION: Pharmaceutical Preparations For
 CC TITLE OF INVENTION: Inhibiting Tumours Associated With Prostate
 CC TITLE OF INVENTION: Adenocarcinoma, Stomach Cancer and Breast Cancer.
 CC NUMBER OF SEQUENCES: 4
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Mr. George Loud
 CC STREET: 2001 Jefferson Davis Highway, Suite 306
 CC CITY: Arlington
 CC STATE: Virginia
 CC COUNTRY: U.S.A.
 CC ZIP: 22202
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patent In Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/07/899,535A
 CC FILING DATE: 16-JUN-1992
 CC CLASSIFICATION: 514
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Loud, George A.
 CC REGISTRATION NUMBER: 25,814
 CC REFERENCE/DOCKET NUMBER: S&B-A835
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 703-415-0960
 CC TELEFAX: 703-415-0962

CC TELEX: 24 8614
 CC INFORMATION FOR SEQ ID NO: 1:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 94 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC HYPOTHETICAL: NO
 CC SEQUENCE 94 AA; 10770 MW; 50094 CN;
 CC
 CC Query Match 40.0%; Score 58; DB 1; Length 94;
 CC Best Local Similarity 37.5%; Pred. No. 5.38e+01;
 CC Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
 CC
 Db 25 KHPINSEWQTDNCETC 40
 YY : || || || || : ||
 QY 153 RSPFDTEWQSKQENTC 168
 CC
 CC RESULT 4
 CC ID US-08-750-359-3 STANDARD; PRT; 14 AA.
 CC XX
 CC AC xxxxxx
 CC DT
 CC DT
 CC DT
 CC DE Sequence 3, Application US/08750359
 CC Sequence 3, Application US/08750359
 CC Patent No. 5916561
 CC GENERAL INFORMATION:
 CC APPLICANT:
 CC TITLE OF INVENTION: Monoclonal antibody against CD44v6
 CC NUMBER OF SEQUENCES: 6
 CC COMPUTER READABLE FORM: disk
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/750,359
 CC FILING DATE:
 CC CLASSIFICATION: 424
 CC INFORMATION FOR SEQ ID NO: 3:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 14 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single stranded
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: Peptide
 CC SEQUENCE 14 AA; 1761 MW; 1187 CN;
 CC
 CC Query Match 38.6%; Score 56; DB 2; Length 14;
 CC Best Local Similarity 50.0%; Pred. No. 8.39e+01;
 CC Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC
 Db 3 FENEWQCKNPPT 14
 YY : || || || :
 QY 156 FDEWQSKQENT 167
 CC
 CC RESULT 5
 CC ID US-08-750-359-5 STANDARD; PRT; 42 AA.
 CC XX
 CC AC xxxxxx
 CC DT
 CC DT
 CC DE Sequence 5, Application US/08750359
 CC Sequence 5, Application US/08750359
 CC Patent No. 5916561

CC GENERAL INFORMATION:
CC APPLICANT: Monoclonal antibody against CD44v6
CC TITLE OF INVENTION: 6
CC NUMBER OF SEQUENCES: 6
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/750,359
CC FILING DATE:
CC CLASSIFICATION: 424

CC INFORMATION FOR SEQ ID NO: 5:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 42 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single stranded

CC TOPOLOGY: linear

CC MOLECULE TYPE: Peptide

CC SEQUENCE 42 AA; 4762 MW; 10496 CN;

Query Match 38.6%; Score 56; DB 2: Length 42;

Best Local Similarity 50.0%; Pred. No. 8.39e+01;

Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 19 FENEWQGNPPT 30

I: |||:|

QY 156 FDEWQSKQENT 167

RESULT 6

ID US-08-359-850-4 STANDARD; PRT; 334 AA.

XX AC xxxxxx

Sequence 4, Application US/08359850

Sequence 4, Application US/08359850

Patent No. 5951982

GENERAL INFORMATION:

APPLICANT: Ziller, Margot

APPLICANT: Herrlich, Peter

APPLICANT: Ponta, Helmut

TITLE OF INVENTION: Use of Antibody-Containing Preparations

TITLE OF INVENTION: for Immunosuppression

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein and Fox

STREET: 1225 Connecticut Avenue

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/359,850

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/963,323

FILING DATE:

APPLICATION NUMBER: DE P 41 34 982.2

FILING DATE: 23-OCT-1991

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 833-7533

TELEFAX: (202) 833-8716

CC INFORMATION FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 334 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

SQ SEQUENCE 334 AA; 36870 MW; 628479 CN;

Query Match 38.6%; Score 56; DB 2: Length 334;

Best Local Similarity 50.0%; Pred. No. 8.39e+01;

Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 139 FENEWQGNPPT 150

I: |||:|

QY 156 FDEWQSKQENT 167

RESULT 7

ID US-07-946-497-5 STANDARD; PRT; 355 AA.

XX AC xxxxxx

Sequence 5, Application US/07946497

Sequence 5, Application US/07946497

Patent No. 5506119

GENERAL INFORMATION:

APPLICANT: HERRLICH, Peter

APPLICANT: PONTA, Helmut

APPLICANT: GUENTHER, Ursula

APPLICANT: MATZKU, Siegfried

APPLICANT: WENZL, Achim

TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA

TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEIN

TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/946,497

FILING DATE: 19921109

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 16915/145

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 355 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: rat protein

SQ SEQUENCE 355 AA; 39023 MW; 704229 CN;

Query Match 38.6%; Score 56; DB 1: Length 355;

Best Local Similarity 50.0%; Pred. No. 8.39e+01;

Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 158 FENEWQGNPPT 169
|:|:|:|
Qy 156 FDEWQSKQENT 167

RESULT 8
ID US-08-478-882-5 STANDARD; PRT: 355 AA.

XX AC xxxxxx

DE Sequence 5, Application US/08478882

XX Sequence 5, Application US/08478882

CC Patent No. 5885575

CC GENERAL INFORMATION:

CC APPLICANT: HERRLICH, Peter

CC APPLICANT: PONTA, Helmut

CC APPLICANT: GUENTHER, Ursula

CC APPLICANT: MATZKU, Siegfried

CC APPLICANT: WENZL, Achim

CC TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA

CC TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,

CC TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY

CC NUMBER OF SEQUENCES: 8

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Foley & Lardner

CC STREET: 3000 K Street, N.W., Suite 500

CC CITY: Washington, D.C.

CC COUNTRY: USA

CC ZIP: 20007-5109

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: PatentIn Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/478.882

CC FILING DATE:

CC CLASSIFICATION: 435

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US/07/946.497

CC FILING DATE: 19921109

CC ATTORNEY/AGENT INFORMATION:

CC NAME: BENT, Stephen A.

CC REGISTRATION NUMBER: 29,768

CC REFERENCE/DOCKET NUMBER: 16915/145

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (202)672-5300

CC TELEFAX: (202)672-5399

CC TELEX: 904136

CC INFORMATION FOR SEQ ID NO: 5:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 355 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC IMMEDIATE SOURCE:

CC CLONE: rat protein

CC SEQUENCE 355 AA; 39023 MW; 704229 CN;

Query Match 38.8%; Score 56; DB 2; Length 355;

Best Local Similarity 50.0%; Pred. No. 8.39e+01;

Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 158 FENEWQGNPPT 169

|:|:|:|

Qy 156 FDEWQSKQENT 167

RESULT 9

ID US-08-483-322-5 STANDARD; PRT: 355 AA.

XX AC xxxxxx

DE Sequence 2, Application US/07946497

XX Sequence 2, Application US/07946497

CC Patent No. 5506119

AC xxxxxx

XX Sequence 5, Application US/08483322

DT Sequence 5, Application US/08483322

XX Patent No. 5760178

DE GENERAL INFORMATION:

XX APPLICANT: HERRLICH, Peter

XX APPLICANT: PONTA, Helmut

XX APPLICANT: GUENTHER, Ursula

XX APPLICANT: MATZKU, Siegfried

XX APPLICANT: WENZL, Achim

XX TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA

XX TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEI

XX TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY

XX NUMBER OF SEQUENCES: 8

XX CORRESPONDENCE ADDRESS:

XX ADDRESSEE: Foley & Lardner

XX STREET: 3000 K Street, N.W., Suite 500

XX CITY: Washington, D.C.

XX COUNTRY: USA

XX ZIP: 20007-5109

XX COMPUTER READABLE FORM:

XX MEDIUM TYPE: Floppy disk

XX COMPUTER: IBM PC compatible

XX OPERATING SYSTEM: PC-DOS/MS-DOS

XX SOFTWARE: PatentIn Release #1.0, Version #1.25

XX CURRENT APPLICATION DATA:

XX APPLICATION NUMBER: US/08/483.322

XX FILING DATE: 07-JUN-1995

XX CLASSIFICATION: 435

XX PRIOR APPLICATION DATA:

XX APPLICATION NUMBER: US 07/946.497

XX FILING DATE: 09-NOV-1992

XX NAME: BENT, Stephen A.

XX ATTORNEY/AGENT INFORMATION:

XX REGISTRATION NUMBER: 29,768

XX REFERENCE/DOCKET NUMBER: 16915/145

XX TELECOMMUNICATION INFORMATION:

XX TELEPHONE: (202)672-5300

XX TELEFAX: (202)672-5399

XX TELEX: 904136

XX INFORMATION FOR SEQ ID NO: 5:

XX SEQUENCE CHARACTERISTICS:

XX LENGTH: 355 amino acids

XX TYPE: amino acid

XX TOPOLOGY: linear

XX IMMEDIATE SOURCE:

XX CLONE: rat protein

XX SEQUENCE 355 AA; 39023 MW; 704229 CN;

Query Match 38.8%; Score 56; DB 1; Length 355;

Best Local Similarity 50.0%; Pred. No. 8.39e+01;

Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 158 FENEWQGNPPT 169

|:|:|:|

Qy 156 FDEWQSKQENT 167

RESULT 10

ID US-07-946-497-2 STANDARD; PRT: 503 AA.

XX AC xxxxxx

XX Sequence 2, Application US/07946497

XX Sequence 2, Application US/07946497

CC Patent No. 5506119

CC	PROTEIN: HCHIM	TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
CC	SEQUENCE: NUMBER	SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEI
CC		TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
CC		TITLE OF INVENTION:

[illegible][illegible]

CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 536 AA; 59333 MW; 1586589 CN;
Query Match 37.9%; Score 55; DB 2; Length 536;
Best Local Similarity 54.5%; Pred. No. 1.05e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 92 YRGCYVTEWDS 102
QY 152 YRSPFDTWQS 162
Search completed: Wed May 10 13:25:04 2000
Job time : 55 secs.

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 92 YRGCYVTEWDS 102
QY 152 YRSPFDTWQS 162
Search completed: Wed May 10 13:25:04 2000
Job time : 55 secs.

RESULT 15
ID US-08-480-150A-4 STANDARD: PRT; 536 AA.
XX
AC xxxxxx
XX
XX
DT
XX
DE Sequence 4, Application US/08480150A
XX
CC Sequence 4, Application US/08480150A
CC Patent No. 5989550
CC GENERAL INFORMATION:
CC APPLICANT: Harris Ph.D., Jeffrey D.
CC APPLICANT: Hsu, Kuang T.
CC APPLICANT: Podolski, Joseph S.
CC TITLE OF INVENTION: Materials and Methods for Immunocontraception
CC NUMBER OF SEQUENCES: 59
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
CC STREET: 6300 Sears Tower, 233 South Wacker Drive
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: United States of America
CC ZIP: 60606-6402
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/480,150A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/149,223
CC FILING DATE: 09-NOV-1993
CC APPLICATION NUMBER: 08/012,990
CC FILING DATE: 29-JAN-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/973,341
CC FILING DATE: 09-NOV-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Clough, David W.
CC REGISTRATION NUMBER: 36,107
CC REFERENCE/DOCKET NUMBER: 31745
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 312/474-6653
CC TELEFAX: 312/474-0448
CC TELEX: 25-3856
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 536 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 536 AA; 59333 MW; 1586589 CN;
Query Match 37.9%; Score 55; DB 2; Length 536;
Best Local Similarity 54.5%; Pred. No. 1.05e+02;

 W P E H

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed May 10 13:23:17 2000; MasPar time 4.59 Seconds
 Tabular output not generated. 195.119 Million cell updates/sec

Title: >US-09-376-430-2
 Description: (151-169) from US09376430A.pep (17 of 25)
 Perfect Score: 145
 Sequence: 1 QVRSPTDTEWQSKQNTCN 19

Scoring table: PAM 150
 Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: pir62
 1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 27.870; Variance 44.016; scale 0.633

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	68	46.9	495	2	cysteine--trna ligase	5.55e-01
2	66	45.5	488	2	deoxyribopyrimidine p	1.22e+00
3	66	45.5	787	2	probable sulfatase (E	1.22e+00
4	64	44.1	431	2	mitochondrial import	2.64e+00
5	63	43.4	222	2	hypothetical protein	3.86e+00
6	62	42.8	201	2	hypothetical protein	5.63e+00
7	62	42.8	230	2	hypothetical protein	5.63e+00
8	62	42.8	479	2	hypothetical protein	5.63e+00
9	62	42.8	1239	2	hypothetical protein	5.63e+00
10	61	42.1	380	2	Ig heavy chain precu	8.18e+00
11	61	42.1	421	2	hypothetical protein	8.18e+00
12	61	42.1	489	1	cytochrome P450 56	8.18e+00
13	61	42.1	584	2	beta-fructofuranosida	8.18e+00
14	60	41.4	376	1	prestalk cathepsin (E	1.18e+01
15	60	41.4	381	2	Ig heavy chain - Raja	1.18e+01
16	60	41.4	1028	2	plasmacytoma-associat	1.18e+01
17	59	40.7	126	2	hypothetical protein	1.70e+01
18	59	40.7	139	2	hypothetical protein	1.70e+01
19	59	40.7	209	2	endo-1,4-beta-xylanas	1.70e+01
20	59	40.7	257	2	probable amino acid A	1.70e+01
21	59	40.7	346	2	molybdenum formylmeth	1.70e+01
22	59	40.7	490	2	probable GTPase/GTP-b	1.70e+01
23	59	40.7	496	2	probable glutathione	1.70e+01

24	59	40.7	529	2	B70211	oligopeptide ABC tran	1.70e+01
25	59	40.7	1777	2	T00490	nonstructural protein	1.70e+01
26	58	40.0	114	2	A34567	beta-microseminoprote	2.45e+01
27	58	40.0	144	2	I64121	DNA-directed DNA poly	2.45e+01
28	58	40.0	320	2	S00935	rlx protein - staphyl	2.45e+01
29	58	40.0	325	2	F64845	probable 2-hydroxyaci	2.45e+01
30	58	40.0	988	4	S58114	transposase (EC 6.5.1	2.45e+01
31	58	40.0	988	1	TQEC26	transposase - Escheri	2.45e+01
32	58	40.0	990	1	TQEC26	transposase - Escheri	2.45e+01
33	58	40.0	1191	2	S76414	beta transducin-like	2.45e+01
34	58	40.0	1615	2	JE0329	low density lipoprote	2.45e+01
35	58	40.0	1615	2	JE0372	low density lipoprote	2.45e+01
36	57	39.3	107	2	A70943	probable repressor pr	3.49e+01
37	57	39.3	172	2	S56905	hypothetical protein	3.49e+01
38	57	39.3	413	1	D71083	probable phosphoester	3.49e+01
39	57	39.3	592	2	S61503	beta-fructofuranosida	3.49e+01
40	57	39.3	592	2	JO0991	beta-fructofuranosida	3.49e+01
41	57	39.3	612	2	A49403	tetratricopeptide-rep	3.49e+01
42	57	39.3	876	2	S20531	hypothetical protein	3.49e+01
43	57	39.3	1028	2	I58164	BIG-1 protein - rat	3.49e+01
44	57	39.3	1146	2	T02766	adhesin WI-1 - Ajello	3.49e+01
45	57	39.3	1703	2	S15047	SNF2 protein - yeast	3.49e+01

ALIGNMENTS

RESULT 1
 ENTRY H70391 #type complete
 TITLE cysteine--trna ligase (EC 6.1.1.16) - Aquifex aeolicus
 ORGANISM #formal_name Aquifex aeolicus
 DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change 16-Jul-1999
 ACCESSIONS H70391
 REFERENCE A70300
 #authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.
 #journal Nature (1998) 392:353-358
 #title The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 #cross-references M01D:98196686
 #accession H70391
 #status preliminary: nucleic acid sequence not shown; translation not shown
 #molecule_type DNA
 #residues 1-495 #label AQF
 #cross-references GB:AE000721; NID:g2983544; PIDN:AAC07125.1; PID:g2983551; GB:AE000657
 #experimental_source strain VF5

GENETICS
 #gene cysS
 CLASSIFICATION #superfamily cysteine--trna ligase
 #keywords aminocyl--trna synthetase; ligase; protein biosynthesis
 #length 495 #molecular-weight 57135 #checksum 5342

Query Match 46.9%; Score 68; DB 2; Length 495;
 Best Local Similarity 46.7%; Pred. No. 5.55e-01;
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

DB 297 HYRSPDLSWERMEE 311

QY 151 QVRSPTDTEWQSKOE 165

RESULT 2
 ENTRY S74829 #type complete
 TITLE deoxyribopyrimidine photolyase - Synechocystis sp. (strain PCC 6803)
 ALTERNATE_NAMES protein slr0854
 ORGANISM #formal_name Synechocystis sp.
 #variety PCC 6803
 DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change

```

22-Jun-1999
S74829
ACCESSIONS
REFERENCE
#authors
Keneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpou,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal
DNA Res. (1996) 3:109-136
#title
Sequence analysis of the genome of the unicellular
cyanobacterium Synechocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references EMBL:D90909; GB:AB001339; NID:g1652844;
#accession MUID:97061201
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-488 #label KAN
#cross-references EMBL:BAAL1790.1; PID:d1018523; PID:g1652872
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
GENETICS
#gene phr
CLASSIFICATION
#superfamily deoxyribodipyrimidine photo-lyase
SUMMARY
#length 488 #molecular-weight 56025 #checksum 8484
Query Match 45.5%; Score 66; DB 2; Length 488;
Best Local Similarity 43.8%; Pred. No. 1.22e+00;
Matches 7; Conservative 5; Mismatches 3; Indels 1; Gaps 1;
Db 319 YRSPQEPFWEENQDH 334
Qy 152 YRSPF-DTEWQSKQEN 166
||||| : : : : :
RESULT 3
ENTRY #type complete
TITLE Probable sulfatase (EC 3.1.1.6.-) atSD - Mycobacterium
ORGANISM tuberculosis (strain H37RV)
#formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
24-Sep-1999
B70535
ACCESSIONS
REFERENCE
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
III, C.E.; Tekaija, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, B.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornaby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession B70535
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-787 #label COL
#cross-references GB:295972; GB:AL123456; NID:g3261790; PID:e319160;
#experimental_source strain H37RV
GENETICS
#gene atsd
KEYWORDS
FEATURE
sulfuric ester hydrolase
89
#modified_site 3-oxoalanine (Cys) #status predicted

22-Jun-1999
S74829
ACCESSIONS
REFERENCE
#authors Keneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpou,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synechocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references EMBL:D90909; GB:AB001339; NID:g1652844;
#accession MUID:97061201
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-488 #label KAN
#cross-references EMBL:BAAL1790.1; PID:d1018523; PID:g1652872
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
GENETICS
#gene phr
CLASSIFICATION
#superfamily deoxyribodipyrimidine photo-lyase
SUMMARY
#length 488 #molecular-weight 56025 #checksum 8484
Query Match 45.5%; Score 66; DB 2; Length 488;
Best Local Similarity 43.8%; Pred. No. 1.22e+00;
Matches 7; Conservative 5; Mismatches 3; Indels 1; Gaps 1;
Db 319 YRSPQEPFWEENQDH 334
Qy 152 YRSPF-DTEWQSKQEN 166
||||| : : : : :
RESULT 3
ENTRY #type complete
TITLE Probable sulfatase (EC 3.1.1.6.-) atSD - Mycobacterium
ORGANISM tuberculosis (strain H37RV)
#formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
24-Sep-1999
B70535
ACCESSIONS
REFERENCE
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
III, C.E.; Tekaija, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, B.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornaby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession B70535
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-787 #label COL
#cross-references GB:295972; GB:AL123456; NID:g3261790; PID:e319160;
#experimental_source strain H37RV
GENETICS
#gene atsd
KEYWORDS
FEATURE
sulfuric ester hydrolase
89
#modified_site 3-oxoalanine (Cys) #status predicted

SUMMARY
#length 787 #molecular-weight 86127 #checksum 6260
Query Match 45.5%; Score 66; DB 2; Length 787;
Best Local Similarity 43.8%; Pred. No. 1.22e+00;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
Db 256 YRGFRDVGWDALREET 271
Qy 152 YRSPFDEWQSKQENT 167
||||| : : : : :
RESULT 4
ENTRY #type complete
TITLE Mitochondrial import protein MPII precursor - yeast
(Saccharomyces cerevisiae)
ALTERNATE_NAMES protein Y13299.09; protein Y11022w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 28-May-1993 #sequence_revision 28-May-1993 #text_change
07-May-1999
ACCESSIONS
REFERENCE
#accession S25196; S49960; S35357
#authors Mearse, A.C.; Blom, J.; Grivell, L.A.; Meijer, M.
#journal EMBO J. (1992) 11:3619-3628
#title MPII, an essential gene encoding a mitochondrial membrane
protein, is possibly involved in protein import into yeast
mitochondria.
#cross-references MUID:93010953
#accession S25196
#molecule_type DNA
#residues 1-431 #label MAA
#cross-references EMBL:X67276; NID:g3966; PID:g3967
REFERENCE
#authors Skelton, J.; Churcher, C.
#submission submitted to the EMBL Data Library, December 1994
#accession S49960
#molecule_type DNA
#residues 1-431 #label SKE
#cross-references EMBL:246881; NID:g599967; PID:g599977; MIPS:Y11022w
REFERENCE
#authors Horst, M.; Jenoe, P.; Kronidou, N.G.; Bolliger, L.; Oppliger,
W.; Scherer, P.; Manning-Krieg, U.; Jascur, T.; Schatz, G.
#journal EMBO J. (1993) 12:3035-3041
#title Protein import into yeast mitochondria: the inner membrane
import site protein ISP45 is the MPII gene product.
#cross-references MUID:93345448
#accession S35357
#molecule_type protein
#residues 82-89; 112-122; 221-228; 295-306; 316-327; 356-370 #label
HOR
GENETICS
#gene SGD:TIM44; MPII
#map_position 9L
#cross-references MIPS:Y11022w; SGD:S0001284
#genome nuclear
KEYWORDS
#length 431 #molecular-weight 48854 #checksum 2985
#molecule_type protein; mitochondrion
SUMMARY
Query Match 44.1%; Score 64; DB 2; Length 431;
Best Local Similarity 42.9%; Pred. No. 2.64e+00;
Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
Db 54 FRDTFKKEWKSQKE 67
Qy 152 YRSPFDEWQSKQKE 165
||||| : : : : :
RESULT 5
ENTRY #type complete
TITLE Hypothetical protein YOL135c - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES hypothetical protein AOE222; hypothetical protein O0505
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change

```



```

ACCESSIONS      26-Feb-1998
REFERENCE        S66832; S72032
#authors        Arino, J.; Casamayor, A.; Camo, F.J.; Gancedo, C.; Lafuente,
#submission     M.J.; Aldea, M.; Casas, C.; Herrero, E.
#accession      Submitted to the Protein Sequence Database, July 1996
#molecule_type DNA
#residues       1-222 ##label API
#cross-references EMBL:Z74877; NID:g1420025; PID:e252305; PID:g1420026;
#experimental_source strain S288C
#authors        Aldea, M.; Piedrafita, L.; Casas, C.; Casamayor, A.; Khalid,
#journal        H.; Balcells, L.; Arino, J.; Herrero, E.
#title          Sequence analysis of a 12 801 bp fragment of the left arm of
#               yeast chromosome XV containing a putative
#               6-phosphofructo-2-kinase gene, a gene for a possible
#               glycopospholipid-anchored surface protein and six other
#               open reading frames.
#cross-references MUID:97051593
#accession      S72032
#status         nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues       1-222 ##label ALD
#cross-references EMBL:X95465; NID:g1628437; PID:e220445; PID:g1628440
#note           the nucleotide sequence was submitted to the EMBL Data
#               Library, January 1996
GENETICS
#map_position 15L
#note          YOL135C
SUMMARY
#length 222 #molecular-weight 25585 #checksum 7876
Query Match      43.4%; Score 63; DB 2; Length 222;
Best Local Similarity 46.7%; Pred. No. 3.86e+00;
Matches          7; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
Db               75 QYRA-FGSIWQVKDQ 88
|||: |||: |||:
QY              151 QYRSPFDTEWQSKOE 165

RESULT          6
ENTRY          T01569 #type complete
TITLE          hypothetical protein A_TM018A10.11 - Arabidopsis thaliana
ORGANISM       #formal_name Arabidopsis thaliana #common_name mouse-ear
#accession     cross
DATE           19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change
#accessions    T01569
REFERENCE      214348
#authors       Dempsey, S.; Harper, M.
#submission    Submitted to the EMBL Data Library, July 1997
#description    The sequence of A. thaliana TM018A10.
#accession     T01569
#status        preliminary; translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues      1-201 ##label DEM
#cross-references EMBL:AF013294; NID:g2252848; PID:g2252858
#experimental_source cultivar Columbia
GENETICS
#map_position 4
#note          A_TM018A10.11
SUMMARY
#length 201 #molecular-weight 23876 #checksum 9489
Query Match      42.8%; Score 62; DB 2; Length 201;
Best Local Similarity 50.0%; Pred. No. 5.63e+00;
Matches          4; Mismatches 3; Indels 1; Gaps 1;
Db               14 SPFNLDWGSMDASC 29
|||: |||: |||: |||:
QY              154 SPFD-TEWQSKQENTC 168

```

```

RESULT          7
ENTRY          C64060 #type complete
TITLE          hypothetical protein HI0296 - Haemophilus influenzae (strain
#accession     Rd KW20)
ORGANISM       #formal_name Haemophilus influenzae
DATE           18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
#accessions    C64060
REFERENCE      A64000
#authors       Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
#               Kirkness, E.F.; Kierlavage, A.R.; Bult, C.J.; Tomb, J.F.;
#               Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
#               FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
#               Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
#               J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
#               M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
#               D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
#               J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
#               Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
#journal        Science (1995) 269:496-512
#title          Whole-genome random sequencing and assembly of Haemophilus
#               influenzae Rd.
#cross-references MUID:95350630
#accession     C64060
#status         nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues      1-230 ##label TIGR
#cross-references GB:U32715; GB:L42023; NID:g1573254; PID:AA21960.1;
#               PID:g1573264; TIGR:HI0296
#superfamily hypothetical protein HI0296
CLASSIFICATION #length 230 #molecular-weight 26789 #checksum 8831
SUMMARY
Query Match      42.8%; Score 62; DB 2; Length 230;
Best Local Similarity 44.4%; Pred. No. 5.63e+00;
Matches          8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
Db               42 QNRSPFPQFASIQKKC 59
|||: |||: |||: |||:
QY              151 QYRSPFDTEWQSKQENTC 168

RESULT          8
ENTRY          T15065 #type complete
TITLE          hypothetical protein R02D3.7 - Caenorhabditis elegans
ORGANISM       #formal_name Caenorhabditis elegans
#accession     20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
#accessions    T15065
REFERENCE      Z18284
#authors       Wu, X.; Antoniou, B.
#submission    Submitted to the EMBL Data Library, December 1997
#description    The sequence of C. elegans cosmid R02D3.
#accession     T15065
#status        preliminary; translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues      1-479 ##label WUX
#cross-references EMBL:AF038615; NID:g2736322; PID:g2736328;
#               PID:AA94147.1; GSPDB:GN00022; CESP:R02D3.7
#experimental_source strain Bristol N2; clone R02D3
GENETICS
#gene          CESP:R02D3.7
#map_position 4
#note          85/3; 162/2; 232/3; 289/3; 363/2; 430/3
SUMMARY
#length 479 #molecular-weight 55372 #checksum 4686
Query Match      42.8%; Score 62; DB 2; Length 479;
Best Local Similarity 47.4%; Pred. No. 5.63e+00;
Matches          9; Conservative 3; Mismatches 5; Indels 2; Gaps 2;
Db               226 QSRDQF-KWFDQROEDTC 243
|||: |||: |||: |||: |||: |||: |||: |||: |||:

```

```

QY 151 QYRSPDTEW-QSKQENTC 168

RESULT 9
ENTRY #type complete
TITLE hypothetical protein sl10068 - Synecocystis sp. (strain PCC
ORGANISM #formal_name Synecocystis sp.
#variety
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
ACCESSIONS PCC 6803
REFERENCE S74355
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugiura, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Nariu, K.; Okumura, S.; Shimpō,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synecocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references EMBL:97061201
#accession S74355
#status preliminary
#molecule_type DNA
#residues 1-1239 #label KAN
#cross-references EMBL:D64001; GB:AB001339; NID:g1001102; PID:d1010924;
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
SUMMARY #length 1239 #molecular-weight 137824 #checksum 4614
Query Match 42.8%; Score 62; DB 2; Length 1239;
Best Local Similarity 72.7%; Pred. No. 5.63e+00;
Matches 8; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
Db 480 QYRI-FDPEWQ 489
QY 151 QYRSPDTEWQ 161

RESULT 10
ENTRY #type complete
TITLE Ig heavy chain precursor - little skate
ORGANISM #formal_name Raja erinacea #common_name little skate
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
09-Sep-1997
ACCESSIONS S12839
REFERENCE S12838
#authors Harding, F.A.; Amemiya, C.T.; Litman, R.T.; Cohen, N.;
Litman, G.W.
#journal Nucleic Acids Res. (1990) 18:6369-6376
#title Two distinct immunoglobulin heavy chain isotypes in a
primitive, cartilaginous fish, Raja erinacea.
#cross-references EMBL:91057118
#accession S12839
#molecule_type mRNA
#residues 1-380 #label HAR
#cross-references EMBL:M29672; NID:g213760; PID:g213761
KEYWORDS heterotrimer; immunoglobulin
SUMMARY #length 380 #molecular-weight 41696 #checksum 3129
Query Match 42.1%; Score 61; DB 2; Length 380;
Best Local Similarity 50.0%; Pred. No. 8.18e+00;
Matches 8; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
Db 305 PLANVEEWNKQYTC 320
QY 155 PFDTQWOSKQENTCN 169

QY 151 QYRSPDTEW-QSKQENTC 168

RESULT 11
ENTRY #type complete
TITLE hypothetical protein F9D16.260 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
23-Jul-1999
ACCESSIONS T05611
REFERENCE Z15419
#authors Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.;
Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
#submision submitted to the Protein Sequence Database, February 1999
#accession T05611
#molecule_type DNA
#residues 1-421 #label BEV
#cross-references EMBL:AL035394
#experimental_source cultivar Columbia; BAC clone F9D16
GENETICS
#map_position 4
#introns 64/1; 161/3
#note F9D16.260
SUMMARY #length 421 #molecular-weight 48579 #checksum 5737
Query Match 42.1%; Score 61; DB 2; Length 421;
Best Local Similarity 44.4%; Pred. No. 8.18e+00;
Matches 8; Conservative 4; Mismatches 5; Indels 1; Gaps 1;
Db 382 QFR-PFDKNAKVQND 398
QY 151 QYRSPDTEWQSKQENTC 168

RESULT 12
ENTRY #type complete
TITLE cytochrome P450 56 - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein YDR402c; spore wall maturation protein DIT2
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
10-Sep-1999
ACCESSIONS B36395; S14228; S69684; S13502
REFERENCE A36395
#authors Briza, P.; Breitenbach, M.; Ellinger, A.; Segall, J.
#journal Genes Dev. (1990) 4:1775-1789
#title Isolation of two developmentally regulated genes involved in
spore wall maturation in Saccharomyces cerevisiae.
#cross-references MIM:91065523
#accession B36395
#molecule_type DNA
#residues 1-489 #label BRI
#cross-references EMBL:X55713
REFERENCE S14228
#authors Segall, J.
#submision submitted to the EMBL Data Library, October 1990
#accession S14228
#molecule_type DNA
#residues 1-444, 'I', 446-489 #label SEG
#cross-references EMBL:X55713; NID:g3655; PIDN:CAA39246.1; PID:g3656
REFERENCE S69684
#authors Dietrich, F.S.
#submision submitted to the EMBL Data Library, July 1995
#description The sequence of S. cerevisiae cosmid 9481, 9509, 9926, 9461,
and lambda 3641.
#accession S69684
#molecule_type DNA
#residues 1-489 #label DIE
#cross-references EMBL:U32274; NID:g927313; PIDN:AAB64842.1;
GENETICS
#gene DIT2; CYP56
#map_position 4R
FUNCTION

```

```
#description      probably involved in dityrosine biosynthesis
CLASSIFICATION   #superfamily yeast cytochrome P450 CYP56; cytochrome P450
homology
KEYWORDS         chromoprotein; electron transfer; heme; iron; monooxygenase;
oxidoreductase; transmembrane protein
FEATURE
8-24            #domain transmembrane #status predicted #label TM1\
33-49           #domain transmembrane #status predicted #label TM2\
435             #binding_site heme iron (Cys) (axial ligand) #status
predicted
SUMMARY          #length 489 #molecular-weight 56071 #checksum 3707
Query Match      42.1%; Score 61; DB 1; Length 489;
Best Local Similarity 46.2%; Pred. No. 8.18e+00;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
Db 455 QFRWSLDPWEK 467
QY 151 QYRSPFTEWQSK 163
I::: I::: I
#cross-references MUID:87172698
#accession A25439
#molecule_type DNA
#residues 1-376 #label DAT
#cross-references GB:M16039; NID:g167859; PIDN:AAA33240.1; PID:g167860
REFERENCE A93604
#authors Pears, C.J.; Mahbubani, H.M.; Williams, J.G.
#journal Nucleic Acids Res. (1985) 13:8853-8866
#title Characterization of two highly diverged but developmentally
co-regulated cysteine proteinase genes in Dictyostelium
discoideum.
#cross-references MUID:86093683
#accession A24110
#molecule_type mRNA
#residues 1-376 #label PEA
#cross-references GB:X03344; NID:g7240; PIDN:CAA27050.1; PID:e300605;
PID:g1834417
GENETICS
#introns *146/1; 194/2; 356/1; 364/2
CLASSIFICATION #superfamily papain
KEYWORDS CAMP binding; cysteine proteinase; glycoprotein; hydrolase
FEATURE
1-18 #domain signal sequence #status predicted #label SIG\
19-121 #domain propeptide #status predicted #label PRO\
122-376 #product prestalk cathepsin #status predicted #label
MAT\
144-187,178-221, #disulfide_bonds #status predicted\
279-365 #active_site Cys, His, Asn #status predicted\
147,286,343 #binding_site carbohydrate (Asn) (covalent) #status
224,237 predicted
SUMMARY #length 376 #molecular-weight 41851 #checksum 2639
Query Match 41.4%; Score 60; DB 1; Length 376;
Best Local Similarity 70.0%; Pred. No. 1.18e+01;
Matches 7; Conservative 2; Mismatches 0; Indels 1; Gaps 1;
Db 31 QYRTAF-TFW 39
QY 151 QYRSPFTEW 160
I::: I::: I
RESULT 15
ENTRY I51174 #type complete
TITLE Ig heavy chain - Raja eglantheria
ORGANISM #formal_name Raja eglantheria
DATE 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change
ACCESSIONS I51174
REFERENCE #authors Anderson, M.; Amemiya, C.; Luer, C.; Litman, R.; Rast, J.;
Nimura, Y.; Litman, G.
#journal Int. Immunol. (1994) 6:1661-1670
#title Complete genomic sequence and patterns of transcription of a
member of an unusual family of closely related,
chromosomally dispersed Ig gene clusters in Raja.
#cross-references MUID:95169645
#accession I51174
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-381 #label AND
#cross-references EMBL:U08010; NID:g727229; PID:g727230
SUMMARY #length 381 #molecular-weight 41612 #checksum 1150
Query Match 41.4%; Score 60; DB 2; Length 381;
Best Local Similarity 63.6%; Pred. No. 1.18e+01;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 314 EWTNKOYTC 324
QY 159 EWTNKOYTC 169
I::: I::: I
#description      probably involved in dityrosine biosynthesis
CLASSIFICATION   #superfamily yeast cytochrome P450 CYP56; cytochrome P450
homology
KEYWORDS         chromoprotein; electron transfer; heme; iron; monooxygenase;
oxidoreductase; transmembrane protein
FEATURE
8-24            #domain transmembrane #status predicted #label TM1\
33-49           #domain transmembrane #status predicted #label TM2\
435             #binding_site heme iron (Cys) (axial ligand) #status
predicted
SUMMARY          #length 489 #molecular-weight 56071 #checksum 3707
Query Match      42.1%; Score 61; DB 1; Length 489;
Best Local Similarity 46.2%; Pred. No. 8.18e+00;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
Db 455 QFRWSLDPWEK 467
QY 151 QYRSPFTEWQSK 163
I::: I::: I
#cross-references MUID:87172698
#accession A25439
#molecule_type DNA
#residues 1-376 #label DAT
#cross-references GB:M16039; NID:g167859; PIDN:AAA33240.1; PID:g167860
REFERENCE A93604
#authors Pears, C.J.; Mahbubani, H.M.; Williams, J.G.
#journal Nucleic Acids Res. (1985) 13:8853-8866
#title Characterization of two highly diverged but developmentally
co-regulated cysteine proteinase genes in Dictyostelium
discoideum.
#cross-references MUID:86093683
#accession A24110
#molecule_type mRNA
#residues 1-376 #label PEA
#cross-references GB:X03344; NID:g7240; PIDN:CAA27050.1; PID:e300605;
PID:g1834417
GENETICS
#introns *146/1; 194/2; 356/1; 364/2
CLASSIFICATION #superfamily papain
KEYWORDS CAMP binding; cysteine proteinase; glycoprotein; hydrolase
FEATURE
1-18 #domain signal sequence #status predicted #label SIG\
19-121 #domain propeptide #status predicted #label PRO\
122-376 #product prestalk cathepsin #status predicted #label
MAT\
144-187,178-221, #disulfide_bonds #status predicted\
279-365 #active_site Cys, His, Asn #status predicted\
147,286,343 #binding_site carbohydrate (Asn) (covalent) #status
224,237 predicted
SUMMARY #length 376 #molecular-weight 41851 #checksum 2639
Query Match 41.4%; Score 60; DB 1; Length 376;
Best Local Similarity 70.0%; Pred. No. 1.18e+01;
Matches 7; Conservative 2; Mismatches 0; Indels 1; Gaps 1;
Db 31 QYRTAF-TFW 39
QY 151 QYRSPFTEW 160
I::: I::: I
RESULT 15
ENTRY I51174 #type complete
TITLE Ig heavy chain - Raja eglantheria
ORGANISM #formal_name Raja eglantheria
DATE 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change
ACCESSIONS I51174
REFERENCE #authors Anderson, M.; Amemiya, C.; Luer, C.; Litman, R.; Rast, J.;
Nimura, Y.; Litman, G.
#journal Int. Immunol. (1994) 6:1661-1670
#title Complete genomic sequence and patterns of transcription of a
member of an unusual family of closely related,
chromosomally dispersed Ig gene clusters in Raja.
#cross-references MUID:95169645
#accession I51174
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-381 #label AND
#cross-references EMBL:U08010; NID:g727229; PID:g727230
SUMMARY #length 381 #molecular-weight 41612 #checksum 1150
Query Match 42.1%; Score 61; DB 2; Length 584;
Best Local Similarity 42.9%; Pred. No. 8.18e+00;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Db 425 FDTAQWTDAGLCK 438
QY 156 FDTAQWTDAGLCK 169
I::: I::: I
RESULT 14
ENTRY KHDOP #type complete
TITLE 'prestalk cathepsin (EC 3.4.22.-) precursor - slime mold
(Dictyostelium discoideum)
ALTERNATE_NAMES cysteine proteinase 2
ORGANISM #formal_name Dictyostelium discoideum
DATE 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change
ACCESSIONS A25439; A24110
REFERENCE A93090
#authors Datta, S.; Firtel, R.A.
#journal Mol. Cell. Biol. (1987) 7:149-159
#title Identification of the sequences controlling cyclic AMP
regulation and cell-type-specific expression of a
prestalk-specific gene in Dictyostelium discoideum.
```

Thu May 11 06:50:03 2000

US-09-376-430-2-17.rpr

Page 6

Search completed: Wed May 10 13:23:26 2000
Job time : 9 secs.

MPRELH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 13:17:02 2000; MasPar time 90.20 Seconds
Tabular output not generated. 6.415 Million cell updates/sec

Title: >US-09-376-430-2
Description: (151-169) from US09376430A.pep (17 of 25)
Perfect Score: 145
Sequence: 1 QYRSPDTEWQSKQENTCN 19

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 28.538; Variance 41.054; scale 0.695

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	68	46.9	495	1 SVC_AQUAE	CYSTEINYL-TRNA SYNTHET	1.89e+1
2	66	45.5	424	1 CRFC_DICDI	CALRETCULIN PRECURSOR	4.40e+01
3	66	45.5	488	1 PHX_SYNY3	DEOXYRIBODIPYRIMIDINE	4.40e+01
4	64	44.1	431	1 IM44_YEAST	MITOCHONDRIAL IMPORT I	1.01e+00
5	63	43.4	344	1 CYS5_DICDI	CYSTEINE PROTEINASE 5	1.52e+00
6	62	42.8	230	1 LEP3_HAEIN	PUTATIVE TYPE 4 PREPIL	2.28e+00
7	61	42.1	301	1 LEC1_RAT	ASIALOGLYCOPROTEIN REC	3.39e+00
8	61	42.1	489	1 CP56_YEAST	CYTOCHROME P450-DIT2 (3.39e+00
9	61	42.1	971	1 YQ1A_CAEEL	HYPOTHETICAL 111.8 KD	3.39e+00
10	60	41.4	376	1 CYS2_DICDI	CYSTEINE PROTEINASE 2	5.04e+00
11	59	40.7	139	1 YRC2_CAEEL	HYPOTHETICAL 16.1 KD P	7.45e+00
12	58	40.0	114	1 P5SP_HUMAN	PROSTATE SECRETED SEMI	1.10e+01
13	58	40.0	144	1 HO1C_HAEIN	DNA POLYMERASE III, CH	1.10e+01
14	58	40.0	151	1 CYS3_DICDI	CYSTEINE PROTEINASE 3	1.10e+01
15	58	40.0	320	1 RLX1_STAAR	RLX PROTEIN.	1.10e+01
16	58	40.0	325	1 YCDW_ECOLI	PUTATIVE 2-HYDROXYACID	1.10e+01
17	58	40.0	389	1 FUT7_MOUSE	ALPHA-(1,3)-FUCOSYLTRA	1.10e+01
18	58	40.0	432	1 IF36_CAEEL	PROBABILE EUKARYOTIC TR	1.10e+01
19	58	40.0	442	1 CYS4_DICDI	CYSTEINE PROTEINASE 4	1.10e+01
20	58	40.0	988	1 TR99_ECOLI	TRANSPPOSASE FOR TRANSP	1.10e+01
21	58	40.0	990	1 TR77_ECOLI	TRANSPPOSASE FOR TRANSP	1.10e+01
22	58	40.0	1191	1 Y143_SYNY3	HYPOTHETICAL WD-REPEAT	1.10e+01
23	57	39.3	172	1 YUM4_YEAST	HYPOTHETICAL 20.3 KD P	1.60e+01

24	57	39.3	413	1 SVI_CIOIN	PROBABLE ISOLEUCYL-TRN	1.60e+01
25	57	39.3	511	1 YE1J_CAEEL	HYPOTHETICAL 58.7 KD P	1.60e+01
26	57	39.3	592	1 INV1_DAUCA	BETA-FRUCTOFURANOSIDAS	1.60e+01
27	57	39.3	612	1 PEX5_YEAST	PEROXISOMAL TARGETING	1.60e+01
28	57	39.3	856	1 VPH1_NEUCR	VACUOLAR ATP SYNTHASE	1.60e+01
29	57	39.3	988	1 TR93_ECOLI	PHOSPHATASE FOR TRANSP	1.60e+01
30	57	39.3	1093	1 PI4K_DICDI	PHOSPHATIDYLINOSITOL 4	1.60e+01
31	57	39.3	1359	1 SNH1_YEAST	NUCLEAR PROTEIN STH1/N	1.60e+01
32	57	39.3	1703	1 SNF2_YEAST	TRANSCRIPTION REGULATO	1.60e+01
33	56	38.6	266	1 TERM_BP03	DNA TERMINAL PROTEIN (2.34e+01
34	56	38.6	404	1 CYC2_YEAST	CYTOCHROME C MITOCHOND	2.34e+01
35	56	38.6	465	1 YQ18_CAEEL	HYPOTHETICAL 53.8 KD P	2.34e+01
36	56	38.6	503	1 CD44_RAT	CD44 ANTIGEN PRECURSOR	2.34e+01
37	56	38.6	1033	1 T1R1_ECOLI	TYPE I RESTRICTION ENZ	2.34e+01
38	56	38.6	1447	1 VGL2_CVPRT	E2 GLYCOPROTEIN PRECUR	2.34e+01
39	56	38.6	1447	1 VGL2_CVPRT	E2 GLYCOPROTEIN PRECUR	2.34e+01
40	56	38.6	1447	1 VGL2_CVPRT	E2 GLYCOPROTEIN PRECUR	2.34e+01
41	56	38.6	1449	1 VGL2_CVPRT	E2 GLYCOPROTEIN PRECUR	2.34e+01
42	56	38.6	1449	1 VGL2_CVPRT	E2 GLYCOPROTEIN PRECUR	2.34e+01
43	55	37.9	269	1 SPEE_ARCFU	PROBABLE SPERMIDINE SY	3.38e+01
44	55	37.9	458	1 PSBC_PROHO	PHOTOSYSTEM II 44 KD R	3.38e+01
45	55	37.9	536	1 ZPB_PIG	ZONA PELLUCIDA SPERM-B	3.38e+01

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	495 AA.
ID	-SVC_AQUAE			
AC	067163;			
DT	15-FEB-2000 (Rel. 39, Created)			
DT	15-FEB-2000 (Rel. 39, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	CYSTEINYL-TRNA SYNTHETASE (EC 6.1.1.16) (CYSTEINE-TRNA LIGASE) (CYSR)			
DE	CYS5 OR AQ_1068.			
GN	Aquifex aeolicus.			
OS	Aquificales; Aquificaceae; Aquifex.			
OC	Bacteria; Aquificales; Aquificaceae; Aquifex.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-VF5;			
RX	MEDLINE; 98196666.			
RA	Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R., Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;			
RA	"The complete genome of the hyperthermophilic bacterium Aquifex aeolicus";			
RL	Nature 392:353-358(1998).			
CC	-!- CATALYTIC ACTIVITY: ATP + L-CYSTEINE + TRNA(CYS) = AMP + PYROPHOSPHATE + L-CYSTEINYL-TRNA(CYS).			
CC	-!- SUBUNIT: MONOMER (BY SIMILARITY).			
CC	-!- SUBCELLULAR LOCATION: CYTOPLASMIC.			
CC	-!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY. STRONG, TO METHIONYL-TRNA SYNTHETASE.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; AE000721; AAC07125.1;			
DR	PROSITE; PS00178; AA_TRNA_LIGASE_I; FALSE_NEG.			
DR	PFAM; PF01406; TRNA-synt_Ie; 1.			
DR	PRINTS; PR09383; TRNASYNTHCS.			
KW	Aminoacyl-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.			
FT	SIMILAR 31 41 "HIGH" REGION.			
FT	SIMILAR 266 270 "KMSKS" REGION.			
FT	BINDING 269 269 ATP (BY SIMILARITY).			
SQ	SEQUENCE 495 AA; 57135 MW; A454658B2BBAE84 CRC64;			

Query Match 46.9%; Score 68; DB 1; Length 495;

Best Local Similarity 46.7%; Pred. No. 1.89e-01;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 297 HYRSPDLFSWEKME 311
QY 151 QYRSPFDTEWQSK 165

RESULT 2
ID CRTCDICDI STANDARD; PRT; 424 AA.
AC Q23858;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CALRETICULIN PRECURSOR.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Dictyostelid; Dictyostelium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RA Mueller-Taubenberger A., Gerisch G.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN BINDS CALCIUM. THERE ARE BOTH HIGH AND
CC LOW AFFINITY CALCIUM-BINDING SITES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
CC -!- SIMILARITY: BELONGS TO THE CALRETICULIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U36937; AAB87719.1; -;
DR DICTYDB; DD00042; -;
DR PRINTS; PR00626; CALRETICULIN.
DR PROSITE; PS00804; CALRETICULIN_2; 1.
DR PROSITE; PS00805; CALRETICULIN_REPEAT; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
DR PFAM; PF00262; calreticulin; 1.
DR Endoplasmic reticulum; Calcium-binding; Repeat; Signal.
KW SIGNAL 1 19 POTENTIAL;
FT CHAIN 20 424 CALRETICULIN.
FT DOMAIN 191 257 4 X 12 AA APPROXIMATE REPEATS.
FT REPEAT 191 202 1-1.
FT REPEAT 211 222 1-2.
FT REPEAT 227 238 1-3.
FT REPEAT 246 257 1-4.
FT DOMAIN 260 298 3 X 11 AA APPROXIMATE REPEATS.
FT REPEAT 260 270 2-1.
FT REPEAT 274 284 2-2.
FT REPEAT 288 298 2-3.
FT DISULFID 105 137 BY SIMILARITY.
FT SITE 421 424 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 424 AA; 48350 MW; BAF273694FB6FC37 CRC64;

Query Match 45.5%; Score 66; DB 1; Length 424;
Best Local Similarity 30.8%; Pred. No. 4.40e-01;
Matches 4; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 22 HFKDTEFDNWSR 34
QY 151 QYRSPFDTEWQSK 163

RESULT 3
ID PHR_SINY3 STANDARD; PRT; 488 AA.
AC Q55081;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE DEOXYRIBODIPYRIMIDINE PHOTOLYASE (EC 4.1.99.3) (DNA PHOTOLYASE)
DE (PHOTOREACTIVATING ENZYME).
GN PHR OR SLR0854.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RA Ng W.O., Pakrasi H.B.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97061201.
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT 'Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.';
RL DNA Res. 3:109-136(1996).
CC -!- FUNCTION: THIS ENZYME CATALYZES THE LIGHT-DEPENDENT MONOMERIZATION
CC (300-600 NM) OF CYCLOBUTYL PYRIMIDINE DIMERS (IN CIS-SYN
CC CONFIGURATION), WHICH ARE FORMED BETWEEN ADJACENT BASES ON THE
CC SAME DNA STRAND, UPON EXPOSURE TO ULTRAVIOLET RADIATION (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: CYCLOBUTADIPYRIMIDINE (IN DNA) = 2 PYRIMIDINE
CC RESIDUES (IN DNA).
CC -!- COFACTOR: CONTAINS 2 CHROMOPHORES: A REDUCED FLAVIN (FADH2) AND
CC AN OXIDIZED 8-HYDROXY-5-DEAZAFLAVIN (F420). BOTH CHROMOPHORES
CC ARE BOUND BY NON-COVALENT INTERACTIONS.
CC -!- SIMILARITY: BELONGS TO THE DNA PHOTOLYASE CLASS-1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U51943; AAB81109.1; -;
DR EMBL; D90909; BAAL7790.1; -;
DR HSSP; P05327; LQNF.
DR PRINTS; PR00147; DNAPHOTOLYASE.
DR PROSITE; PS00394; DNA_PHOTOLYASES_1_1; 1.
DR PROSITE; PS00691; DNA_PHOTOLYASES_1_2; 1.
DR PFAM; PF00875; DNA_photolypase; 1.
KW Lyase; Chromophore; Flavoprotein; FAD; DNA repair; DNA-binding.
FT DNA_BIND 333 352 H-T-H MOTIF (POTENTIAL).
SQ SEQUENCE 488 AA; 56025 MW; BEE64ABA0919332 CRC64;
Query Match 45.5%; Score 66; DB 1; Length 488;
Best Local Similarity 43.8%; Pred. No. 4.40e-01;
Matches 7; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

Db 319 YRSPFOEFFWEENQDH 334
QY 152 YRSPFDTEWQSK 166

RESULT 4
ID IM44_YEAST STANDARD; PRT; 431 AA.
AC Q01852;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MITOCHONDRIAL IMPORT INNER MEMBRANE TRANSLOCASE SUBUNIT TIM44
DE PRECURSOR (MITOCHONDRIAL PROTEIN IMPORT PROTEIN 1) (INNER MEMBRANE
DE IMPORT SITE PROTEIN 45) (ISP45) (MEMBRANE IMPORT MACHINERY PROTEIN
DE TIM44).
GN TIM44 OR MP11 OR MIM44 OR ISP45 OR YIL022W.
OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AB320;
 RX MEDLINE; 93010953.
 RA Maarse A.C., Blom J., Grivell L.A., Meijer M.;
 RT "MP1, an essential gene encoding a mitochondrial membrane protein,
 RT is possibly involved in protein import into yeast mitochondria";
 RL EMBO J. 11:3619-3628(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
 RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RX MEDLINE; 93345448.
 RA Horst M., Jenoe P., Kronidou N.G., Bolliger L., Oppliger W.,
 RA Scherer P., Manning-Krieg U., Jascut I., Schatz G.;
 RT "Protein import into yeast mitochondria: the inner membrane import
 RT site protein ISP45 is the MP1 gene product";
 RL EMBO J. 12:3035-3041(1993).
 CC -!- FUNCTION: INVOLVED IN PROTEIN IMPORT INTO THE MITOCHONDRION.
 CC PROBABLY INVOLVED IN TRANSLLOCATION ACROSS THE INNER MEMBRANE.
 CC AS A BINDING PROTEIN REQUIRED FOR DRIVING THE IMPORT OF
 CC PREPROTEINS. RECRUITS MITOCHONDRIAL HSP70 AND ITS CO-CHAPERONE
 CC (MGFI) TO DRIVE PROTEIN TRANSLLOCATION INTO THE MATRIX USING ATP
 CC AS AN ENERGY SOURCE.
 CC -!- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT
 CC LEAST 3 DIFFERENT PROTEINS (TIM17, TIM23, TIM44).
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
 CC -!- SIMILARITY: BELONGS TO THE TIM44 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z46881; CA986970.1; -;
 DR EMBL; X87276; CA947693.1; -;
 DR PIR; S25196; S25196.
 DR SGD; L0001138; MP1.
 KW Mitochondrion; Inner membrane; Transport; Protein transport;
 KW Translocation; Transit peptide; ATP-binding.
 FT TRANSIT 1 ?
 FT CHAIN 1 ? MITOCHONDRION.
 FT NP_BIND 101 108
 FT ATP (POTENTIAL).
 SQ SEQUENCE 431 AA; 48854 MW; 8E98C738178B1E44 CRC64;
 Query Match 44.1%; Score 64; DB 1; Length 431;
 Best Local Similarity 42.9%; Pred. No. 1.01e+00;
 Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 Db 54 FRDTFKKEWKSQE 67
 QY 152 YRSPFDTEWQSKQE 165
 RESULT 5
 ID CY55_DICDI STANDARD; PRT; 344 AA.
 AC P54640;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE CYSTEINE PROTEINASE 5 PRECURSOR (EC 3.4.22.-).
 CPRE OR CP5.
 OS Dictyostelium discoideum (Slime mold).
 CC Eukaryota; Dictyostelid; Dictyostelium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RX MEDLINE; 96081966.
 RA Souza G.M., Hirai J., Mehta D.P., Freeze H.H.;
 RT "Identification of two novel Dictyostelium discoideum cysteine
 RT proteinases that carry N-acetylglucosamine-1-P-modification";
 RL J. Biol. Chem. 270:28938-28945(1995).
 CC -!- SUBCELLULAR LOCATION: LYSOSOMAL.
 CC -!- DEVELOPMENTAL STAGE: PRESENT IN THE VEGETATIVE PHASE AND DECREASES
 CC WITH THE START DEVELOPMENT, REAPPEARS IN LOW LEVELS WHEN THE
 CC FRUITING BODY IS FORMED.
 CC -!- PTM: PHOSPHOGLYCOSYLATED, CONTAINS GLCNAC-ALPHA-1-P-SER RESIDUES.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
 CC PAPAIN FAMILY OF THIOL PROTEASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L36205; AAA92018.1; -;
 DR HSSP; P07711; ICUL.
 DR DICTYDB; DD01061; CPRE.
 DR PRINTS; PR00705; PAPAIN.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
 DR PFAM; PF00112; Peptidase_C1; 1.
 KW Hydrolase; Thiol protease; Lysosome; Zymogen; Glycoprotein;
 KW Phosphorylation; Signal.
 FT SIGNAL 1 17
 FT PROPEP 18 111
 FT CHAIN 112 344
 FT DOMAIN 196 340
 FT ACT_SITE 136 136
 FT ACT_SITE 272 272
 FT ACT_SITE 311 311
 FT ACT_SITE 313 174
 FT DISULFID 167 207
 FT DISULFID 265 333
 FT CARBOHYD 110 110
 FT CARBOHYD 297 297
 SQ SEQUENCE 344 AA; 37212 MW; 82F3F35B8C147BA8 CRC64;
 Query Match 43.4%; Score 63; DB 1; Length 344;
 Best Local Similarity 35.3%; Pred. No. 1.52e+00;
 Matches 6; Conservative 7; Mismatches 3; Indels 1; Gaps 1;
 Db 25 QYRNAF-TDWMITHOKS 40
 QY 151 QYRSPFDTEWQSKQENT 167
 RESULT 6
 ID LEP3_HAEIN STANDARD; PRT; 230 AA.
 AC P44620;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE PUTATIVE TYPE 4 PREPILIN-LIKE PROTEIN SPECIFIC LEADER PEPTIDASE
 DE (EC 3.4.99.-)
 GN HOED OR HOED OR HI0296.
 OS Haemophilus influenzae.
 CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE: 95350630.
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fithugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-L., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: CLEAVES TYPE-4 FIMBRIAL LEADER SEQUENCE AND METHYLATES
CC THE N-TERMINAL (GENERALLY PHE) RESIDUE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (PROBABLE).
CC -!- LEADER PEPTIDASE FAMILY C20; ALSO KNOWN AS TYPE
CC III LEADER PEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32715; AAC21960.1; -;
DR TIGR; HI0296; -;
DR PFAM; PF01478; Peptidase_C20; 1.
KW Hydrolase; Protease; Transport; Transmembrane; Inner membrane.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 60 80 POTENTIAL.
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 114 134 POTENTIAL.
FT TRANSMEM 140 160 POTENTIAL.
FT TRANSMEM 181 201 POTENTIAL.
FT TRANSMEM 208 228 POTENTIAL.
SQ SEQUENCE 230 AA; 26788 MW; CD12ABE9B74818D0 CRC64;
Query Match 42.8%; Score 62; DB 1; Length 230;
Best Local Similarity 44.4%; Pred. No. 2.28e+00;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
Pb 42 QNRSPQPHFASIQKKC 59
QY 151 QYRSPDTEWQSKQNTC 168
RESULT 7
ID LEC1_RAT STANDARD; PRT; 301 AA.
AC P08290;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE ASIALOGLYCOPROTEIN RECEPTOR R2/3 (HEPATIC LECTIN 2/3) (RHL-2) (ASGP-R)
DE (ASGPR).
GN ASGR2 OR ASGR-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX McPhaul M., Berg P.;
RT "Identification and characterization of cDNA clones encoding two
RT homologous proteins that are part of the asialoglycoprotein
RT receptor.";

RL Mol. Cell. Biol. 7:1841-1847(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87250656.
RA Halberg D.F., Wager R.E., Farrell D.C., Hildreth J., Quesenberry M.S.,
RA Loeb J.A., Holland E.C., Drickamer K.;
RT "Major and minor forms of the rat liver asialoglycoprotein receptor
RT are independent galactose-binding proteins. Primary structure and
RT glycosylation heterogeneity of minor receptor forms.";
RL J. Biol. Chem. 262:9828-9838(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=LIVER;
RX MEDLINE: 89170119.
RA Sanford J.P., Elliott R.W., Doyle D.;
RT "Asialoglycoprotein receptor genes are linked on chromosome 11 in the
RT mouse.";
RL DNA 7:721-728(1988).
RN [4]
RP SEQUENCE OF 201-301.
RX MEDLINE: 84111554.
RA Drickamer K., Mamou J.F., Binns G., Leung J.O.;
RT "Primary structure of the rat liver asialoglycoprotein receptor.
RT Structural evidence for multiple polypeptide species.";
RL J. Biol. Chem. 259:770-778(1984).
CC -!- FUNCTION: MEDIATES THE ENDOCYTOSIS OF PLASMA GLYCOPROTEINS TO
CC WHICH THE TERMINAL SIALIC ACID RESIDUE ON THEIR COMPLEX
CC CARBOHYDRATE MOIETIES HAS BEEN REMOVED. THE RECEPTOR RECOGNIZES
CC TERMINAL GALACTOSE AND N-ACETYL GALACTOSAMINE UNITS. AFTER LIGAND
CC BINDING TO THE RECEPTOR, THE RESULTING COMPLEX IS INTERNALIZED AND
CC TRANSPORTED TO A SORTING ORGANELLE, WHERE RECEPTOR AND LIGAND ARE
CC DISSOCIATED. THE RECEPTOR THEN RETURNS TO THE CELL MEMBRANE
CC SURFACE.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN HEPATIC PARENCHYMAL
CC CELLS.
CC -!- MISCELLANEOUS: CALCIUM IS REQUIRED FOR LIGAND BINDING.
CC -!- MISCELLANEOUS: TWO TYPES OF RAT HEPATIC LECTIN HAVE BEEN
CC IDENTIFIED, RHL-1 AND RHL-2/3, HAVING A RELATIVE ABUNDANCE OF 4:1.
CC RHL-2 AND RHL-3 ONLY DIFFERS IN THEIR CARBOHYDRATE STRUCTURES.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16347; AAA42038.1; -;
DR EMBL; J02762; AAA41522.1; -;
DR EMBL; X07636; CAA30476.1; -;
DR PIR; A25417; LNRT2.
DR PIR; A26888; A26888.
DR PIR; A28462; A28462.
DR PIR; A31601; A31601.
DR HSSP; P06734; 1KJE.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_2; 1.
DR PFAM; PF00059; lectin_c; 1.
KW Lectin; Glycoprotein; Receptor; Endocytosis; Transmembrane;
KW Calcium; Signal-anchor; Phosphorylation.
FT DOMAIN 1 58 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 59 79 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 80 301 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 169 295 C-TYPE LECTIN (LONG FORM).
FT DISULFID 170 181 BY SIMILARITY.
FT DISULFID 198 293 BY SIMILARITY.
FT DISULFID 271 285 BY SIMILARITY.
FT CARBOHYD 97 97
FT CARBOHYD 119 119
FT CARBOHYD 165 165


```

FT CONFLICT 445      M -> I (IN REF. 1).
SQ SEQUENCE 489 AA; 56071 MW; 9A301276DDDBE18F CRC64;

Query Match          42.1%; Score 61; DB 1; Length 489;
Best Local Similarity 46.2%; Pred. No. 3.39e+00;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Ddb 455 QFRWLDPEWEK 467
QY 151 QYRSPFDTEWQSK 163
    | : : : | : : |
RESULT 9
ID YQIA_CAEEL STANDARD; PRT; 971 AA.
AC QO9281;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHETICAL 111.8 KD PROTEIN C45G9.10 IN CHROMOSOME III.
OS C45G9.10.
OC Caenorhabditis elegans.
CC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditida;
CC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
SEQUENCE FROM N.A.
RA Bantley D., Waterston R.;
RR STRAIN-BRISTOL N2;
RL Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outpost -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U21323; AAA62553.1; -.
WORMEP; C45G9.10; CE01851.
Hypothetical protein.
SEQUENCE 971 AA; 111772 MW; 6AA5EB10E2B7693 CRC64;

Query Match          42.1%; Score 61; DB 1; Length 971;
Best Local Similarity 53.3%; Pred. No. 3.39e+00;
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

Ddb 674 SVFNWGWE-KOENAC 687
QY 154 SPFDTEWQSKQENTC 168
    | : : | : ||| : |
RESULT 10
ID CYS2_DICDI STANDARD; PRT; 376 AA.
AC P04989;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CYSTEINE PEPTIDINASE 2 PRECURSOR (EC 3.4.22.-) (PRESTALK CATHEPSIN).
OS CPBB OR CP2.
OC Dictyostelium discoideum (Slime mold);
CC Eukaryota; Dictyosteliida; Dictyostelium.
RN [1]
SEQUENCE FROM N.A.
RA Datta S., Firtel R.A.;
RR MEDLINE; 87172698.
RL "Identification of the sequences controlling cyclic AMP regulation
and cell-type-specific expression of a prestalk-specific gene in
Dictyostelium discoideum";
Mol. Cell. Biol. 7:149-159(1987).
[2]
SEQUENCE FROM N.A.
RA Pears C.J., Mahubani H.M., Williams J.G.
RR MEDLINE; 86093683.

```

RT "Characterization of two highly diverged but developmentally co-regulated cysteine proteinase genes in Dictyostelium discoideum.";
RL Nucleic Acids Res. 13:8853-8866(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87057653.
RA Gomer R.H., Datta S., Firtel R.A.;
RT "Cellular and subcellular distribution of a CAMP-regulated prestalk protein and prespore protein in Dictyostelium discoideum: a study on the ontogeny of prestalk and prespore cells.";
RL J. Cell Biol. 103:1999-2015(1986).
CC -1- FUNCTION: CYSTEINE PROTEINASES 1 AND 2 ARE BELIEVED TO PARTICIPATE IN THE BREAKDOWN OF PROTEIN DURING DIFFERENTIATION OF DICTYOSTELIUM AS A RESPONSE TO STARVATION.
CC -1- SUBCELLULAR LOCATION: LYSOSOMAL.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED BY PRESTALK CELLS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE PAPAINE FAMILY OF THIOLESTERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16039; AAA33240.1; -
DR EMBL; X03344; CAA27050.1; -
DR PIR; A25439; KHDOL.
DR HSSP; P07711; ICJL.
DR DICTDB; DD02004; CPRB.
DR PRINTS; PR00705; PAPAINE.
DR PROSITE; PS00139; THIOLESTERASE_CYS; 1.
DR PROSITE; PS00639; THIOLESTERASE_HIS; 1.
DR PROSITE; PS00640; THIOLESTERASE_ASN; 1.
DR PFAM; PF00112; Peptidase_C1; 2.
KW Hydrolase; Thiol protease; Lysosome; zymogen; Signal.
FT SIGNAL 1 18
FT PROPEP 19 122 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 123 376 CYSTEINE PROTEINASE 2.
FT ACT_SITE 147 147 BY SIMILARITY.
FT ACT_SITE 286 286 BY SIMILARITY.
FT ACT_SITE 343 343 BY SIMILARITY.
FT DISULFD 144 187 BY SIMILARITY.
FT DISULFD 178 221 BY SIMILARITY.
FT DISULFD 279 365 BY SIMILARITY.
SQ SEQUENCE 376 AA; 41851 MW; 161B293643456422 CRC64;

Query Match 41.4%; Score 60; DB 1; Length 376;
Best Local Similarity 70.0%; Pred. No. 5.04e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

Db 31 QYRTAF-TFW 39
Qy 151 QYRSPFDTEW 160

RESULT 11
ID YRC2_CAEEL STANDARD; PRT; 139 AA.
AC Q10041;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL 16.1 KD PROTEIN K01A12.2 IN CHROMOSOME X.
GN K01A12.2;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX Geisel C.;

RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U23183; AAA64339.1; -
DR WORMPEP; K01A12.2; CE01987.
KW Hypothetical protein.
SQ SEQUENCE 139 AA; 16141 MW; E02E1BC5ECB96879 CRC64;

Query Match 40.7%; Score 59; DB 1; Length 139;
Best Local Similarity 45.5%; Pred. No. 7.45e+00;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 93 YRTNFDQWRG 103
Qy 152 YRSPFDTEWQS 162

RESULT 12
ID PSSP_HUMAN STANDARD; PRT; 114 AA.
AC P08118; P11999;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PROSTATE SECRETED SEMINAL PLASMA PROTEIN PRECURSOR (PSP-94) (BETA-DE MICROSEMINOPROTEIN) (SEMINAL PLASMA BETA-INHIBIN) (IMMUNOGLOBULIN BINDING FACTOR) (IGBF) (PROSTATE SECRETORY PROTEIN PSP94) (PN44).
GN MSMB OR PRSP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87161231.
RA Mbikay M., Nolet S., Fournier S., Benjannet S., Chapdelaine P., Paradis G., Dube J.Y., Tremblay R., Lazure C., Seidah N.G., Chretien M.;
RT "Molecular cloning and sequence of the cDNA for a 94-amino-acid seminal plasma protein secreted by the human prostate.";
RL DNA 6:23-29(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91274357.
RA Nolet S., Mbikay M., Chretien M.;
RT "Prostatic secretory protein PSP94: gene organization and promoter sequence in rhesus monkey and human.";
RL Biochim. Biophys. Acta 1089:247-249(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90211299.
RA Green C.B., Liu W.Y., Kwok S.C.M.;
RT "Cloning and nucleotide sequence analysis of the human beta-microseminoprotein gene.";
RL Biochem. Biophys. Res. Commun. 167:1184-1190(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90073664.
RA Ulyback M., Lindstrom C., Weiber H., Abrahamsson P.A., Lilja H., Lundwall A.;
RT "Molecular cloning of a small prostate protein, known as beta-microseminoprotein, PSP94 or beta-inhibin, and demonstration of transcripts in non-genital tissues.";
RL Biochem. Biophys. Res. Commun. 164:1310-1315(1989).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE-PROSTATE;
RX MEDLINE; 94115955.


```
RESULT 14
ID CYS3_DICDI STANDARD; PRT; 151 AA.
AC Q23894;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CYSTEINE PROTEINASE 3 (EC 3.4.22.-) (CYSTEINE PROTEINASE II)
DE (FRAGMENT).
GN CPRC OR CP3.
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Dictyostellida; Dictyostellum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 24397;
RA Presse F., Bogdanovsky-Sequeval D., Mathieu M., Fellenbok B.;
RT "Structural analysis of a developmentally regulated sequence encoding
RT for a cysteine proteinase in Dictyostellium discoideum."
RL Mol. Gen. Genet. 203:324-332(1986).
CC -!- SUBCELLULAR LOCATION: LYSOSOMAL (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
CC PAPAINE FAMILY OF THIOL PROTEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03930; CAA27569.1; -
DR HSP; P10056; MEG.
DR DICTYD; D001062; CPRC.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; FALSE_NEG.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; PARTIAL.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; PARTIAL.
DR PFAM; PF00112; Peptidase_C1; 1.
KW Hydrolase; Thiol protease; Lysosome.
FT NON_TER 1
FT ACT_SITE 100 100 BY SIMILARITY.
FT DISULFID 97 139 BY SIMILARITY.
FT NON_TER 151 151
SQ SEQUENCE 151 AA; 17115 MW; 63852EB5D1AAD0A6 CRC64;

Query Match 40.0%; Score 58; DB 1; Length 151;
Best Local Similarity 23.5%; Pred. No. 1.10e+01;
Matches 4; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

Db 73 HFKQPLNVDWREKDAVT 89
::: |: |: |: |
QY 151 QYRSPFDTEWQSKQENT 167

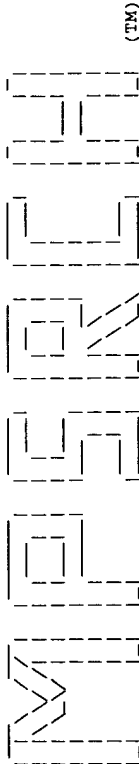
RESULT 15
ID RLX1_STAAR STANDARD; PRT; 320 AA.
AC P12054;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1989 (Rel. 12, Last annotation update)
DE RLX PROTEIN.
DE Staphylococcus aureus.
OS Staphylococcus aureus.
OS Plasmid pS194.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE; 88189810.
RA Projan S.J., Moghazeh S., Novick R.P.;
RT "Nucleotide sequence of pS194, a streptomycin-resistance plasmid from
RT Staphylococcus aureus."
RL Nucleic Acids Res. 16:2179-2187(1988).
CC -!- FUNCTION: THIS PROTEIN IS PROBABLY REQUIRED FOR RELAXATION
CC COMPLEX FORMATION AND PLASMID MOBILIZATION BY CONJUGATIVE
```

```
-----
PLASMIDS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06627; CAA29840.1; -
DR PIR; S00935; S00935.
KW Plasmid.
SQ SEQUENCE 320 AA; 36997 MW; 256691BFED6DFC4 CRC64;

Query Match 40.0%; Score 58; DB 1; Length 320;
Best Local Similarity 29.4%; Pred. No. 1.10e+01;
Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 252 EYADTFEVDWDAFAENS 268
::: |: |: |: |
QY 151 QYRSPFDTEWQSKQENT 167

Search completed: Wed May 10 13:18:41 2000
Job time : 99 secs.
```



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 12:27:36 2000; MasPar time 91.56 Seconds
3.991 Million cell updates/sec
Tabular output not generated.

Title: >US-09-376-430-2
Description: (62-73) from US09376430A.pep (11 of 25)
Perfect Score: 87
Sequence: 1 FNGDEAYDQCTN 12

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss:prot38
1:swissprot

Statistics: Mean 24.606; Variance 28.651; scale 0.859

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	55	63.2	663	1	CG2_CHICK 72 KD TYPE IV COLLAGEN	6.13e-01
2	55	63.2	1380	1	ZMS1_YEAST ZINC FINGER PROTEIN ZM	6.13e-01
3	52	59.8	2242	1	PYR1_SQUAC CAD PROTEIN [INCLUDES:	2.71e+00
4	51	58.6	350	1	KC22_BOVIN CASEIN KINASE II, ALPH	4.39e+00
5	51	58.6	350	1	KC22_HUMAN CASEIN KINASE II, ALPH	4.39e+00
6	51	58.6	350	1	KC22_CHICK CASEIN KINASE II, ALPH	4.39e+00
7	51	58.6	649	1	SCAG_HUMAN AMILORIDE-SENSITIVE SO	4.39e+00
8	51	58.6	650	1	SCAG_RAT AMILORIDE-SENSITIVE SO	4.39e+00
9	51	58.6	660	1	CG2_HUMAN 72 KD TYPE IV COLLAGEN	4.39e+00
10	50	57.5	335	1	KC2A_DROME CASEIN KINASE II, ALPH	7.05e+00
11	50	57.5	350	1	KC22_XENLA CASEIN KINASE II, ALPH	7.05e+00
12	50	57.5	356	1	GBA1_DICDI GUANINE NUCLEOTIDE-BIN	7.05e+00
13	50	57.5	360	1	KC2A_CAEEL CASEIN KINASE II, ALPH	7.05e+00
14	50	57.5	391	1	KC21_MOUSE CASEIN KINASE II, ALPH	7.05e+00
15	50	57.5	391	1	KC21_RABIT CASEIN KINASE II, ALPH	7.05e+00
16	50	57.5	391	1	KC21_RAT CASEIN KINASE II, ALPH	7.05e+00
17	50	57.5	391	1	KC21_CHICK CASEIN KINASE II, ALPH	7.05e+00
18	50	57.5	391	1	KC21_HUMAN CASEIN KINASE II, ALPH	7.05e+00
19	50	57.5	398	1	CAR3_CANAL CANDIDAPEPIN 3 PRECUR	7.05e+00
20	50	57.5	551	1	REFR_VACCV RIFAMPICIN RESISTANCE	7.05e+00
21	50	57.5	551	1	REFR_VARY RIFAMPICIN RESISTANCE	7.05e+00
22	50	57.5	630	1	YB5_YEAST HYPOTHETICAL 71.9 KD P	7.05e+00
23	50	57.5	992	1	VP41_YEAST VACUOLAR ASSEMBLY PROT	7.05e+00

24	49	56.3	355	1	CG2_CHICK STANDARD; PRT: 663 AA.	
25	49	56.3	3255	1	O90611;	
26	49	56.3	3255	1	01-NOV-1997 (Rel. 35, Created)	
27	48	55.2	122	1	01-NOV-1997 (Rel. 35, Last sequence update)	
28	48	55.2	189	1	72 KD TYPE IV COLLAGENASE PRECURSOR (EC 3.4.24.24)	
29	48	55.2	411	1	(MATRIX METALLOPROTEINASE-2) (MMP-2) (GELATINASE A).	
30	48	55.2	427	1	MMP2.	
31	48	55.2	561	1	Gallus gallus (Chicken).	
32	48	55.2	700	1	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinoptera; Euteleostomi; Galliformes; Phasianidae; Phasianinae; Gallus	
33	48	55.2	803	1	YDHV_YEAST PUTATIVE EARLY 40.3 KD	1.13e+01
34	48	55.2	948	1	DIM_FUJOSE POLYPROTEIN [CO	1.13e+01
35	48	55.2	1058	1	PEX6_SCOSOMAL BISPHOSPHATE	1.78e+01
36	48	55.2	1102	1	YMS6_YEAST SPECIFIC P	1.78e+01
37	48	55.2	1763	1	PKSM_BACSTRATION PROTEIN P	1.78e+01
38	48	55.2	2109	1	PGCA_CHICK 77.9 KD P	1.78e+01
39	48	55.2	4427	1	PKSL_BACSU PROTEIN	1.78e+01
40	47	54.0	184	1	YAD2_CLOAB YAD2 P	1.78e+01
41	47	54.0	321	1	YHOG_KLEPN Y-6.	1.78e+01
42	47	54.0	373	1	VGLX_RSVEX Geronte	1.78e+01
43	47	54.0	509	1	KPYK_SCHPO PAKD	1.78e+01
44	47	54.0	622	1	KFE1_YEAST PAKD	1.78e+01
45	47	54.0	662	1	COG2_RABIT 72 K	1.78e+01

ALIGNMENTS

RESULT 1			
CG2_CHICK	STANDARD;	PRT: 663 AA.	
Q90611;			
01-NOV-1997 (Rel. 35, Created)			
01-NOV-1997 (Rel. 35, Last sequence update)			
01-NOV-1997 (Rel. 35, Last annotation update)			
72 KD TYPE IV COLLAGENASE PRECURSOR (EC 3.4.24.24)			
(MATRIX METALLOPROTEINASE-2) (MMP-2) (GELATINASE A).			
MMP2.			
Gallus gallus (Chicken).			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinoptera; Euteleostomi; Galliformes; Phasianidae; Phasianinae; Gallus			
Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus			
[1]			
SEQUENCE FROM N.A.			
TISSUE=EMBRYO;			
MEDLINE; 94280397.			
Aimes R.T., French D.L., Quigley J.P.;			
"Cloning of a 72 kDa matrix metalloproteinase (gelatinase) from chicken embryo fibroblasts using gene family PCR: expression of the gelatinase increases upon malignant transformation.";			
Biochem. J. 300:729-736(1994).			
[2]			
SEQUENCE OF 27-41 AND 107-122.			
MEDLINE; 91161603.			
Chen J.-M., Aimes R.T., Ward G.R., Youngleib G.L., Quigley J.P.;			
"Isolation and characterization of a 70-kDa metalloproteinase (gelatinase) that is elevated in Rous sarcoma virus-transformed chicken embryo fibroblasts";			
J. Biol. Chem. 266:5113-5121(1991).			
[-] CATALYTIC ACTIVITY: CLEAVAGE OF GELATIN TYPE I AND COLLAGEN TYPES IV, V, VII, X. CLEAVES THE COLLAGEN-LIKE SEQUENCE PRO-GLN-GLY-[-] ILE-ALA-GLY-GLN.			
[-] COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY.			
[-] TISSUE SPECIFICITY: PRODUCED BY NORMAL SKIN FIBROBLASTS.			
[-] SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.			
[-] SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE II-LIKE DOMAINS.			
[-] SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A (ZINC METALLOPROTEASE) ALSO KNOWN AS MATRXIN SUBFAMILY.			
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			

```
DR EMBL: U07775; AAA19596.1; -.
DR HSSP: P08253; INTG.
DR PRINTS; PR00013; FNTYPEII.
DR PRINTS; PR00138; MATRXIN.
DR PROSITE; PS00023; FIBONECTIN.2; 3.
DR PROSITE; PS00024; HEMOXIN; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00346; CYSTEINE_SWITCH; 1.
DR PFAM; PF00040; fn2; 3.
DR PFAM; PF00045; hemopexin; 4.
DR PFAM; PF00413; peptidase_M10; 1.
KW Hydrolase; Metalloprotease; Zinc; Zymogen; Calcium;
KW Collagen degradation; Extracellular matrix; Signal.
FT SIGNAL 1 26
FT PROPEP 27 106
FT CHAIN 107 663
FT DOMAIN 97 103
FT DOMAIN 107 218
FT DOMAIN 219 393
FT DOMAIN 394 468
FT DOMAIN 469 663
FT DOMAIN 223 280
FT DOMAIN 281 338
FT DOMAIN 339 396
FT METAL 400 400
FT ACT_SITE 401 401
FT METAL 404 404
FT METAL 410 410
FT DISULFID 472 663
FT CONFLICT 40 40
FT CONFLICT 116 116
FT CONFLICT 122 122
FT SEQUENCE 663 AA; 74941 MW; 8D6FDA4E67C3EBCA CRC64;

Query Match 63.2%; Score 55; DB 1; Length 663;
Best Local Similarity 63.6%; Pred. No. 6.13e-01;
Matches 7; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 294 F0G-QSDQCT 303
QY 62 FNGDEAYDQCT 72

RESULT 2
ID ZMS1_YEAST STANDARD; PRT; 1380 AA.
AC P46974;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ZINC FINGER PROTEIN ZMS1.
GN ZMS1 OR YJR127C OR J2052.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RC11-6A;
RA Thomas D., Barbey R., Surdin-Kerjan Y.;
RL Submitted (XXX-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA ROSE M., Koetter P., Entian K.D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -!- CELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -!- SIMILARITY: STRONG, TO YEAST YML081W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

CC DR EMBL: L26506; AAA35240.1; -.
CC DR EMBL: Z49627; CAA89658.1; -.
CC DR HSSP: P07248; IARD.
CC DR SGD; L0002569; ZMS1.
CC DR PRINTS; PR00048; ZINC_FINGER.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2; 2.
CC DR PFAM; PF00096; zf-C2H2; 2.
KW DNA-binding; Nuclear protein; Zinc-finger; Metal-binding.
FT DOMAIN 64 67
FT ZN_FING 151 173
FT ZN_FING 179 202
FT DOMAIN 424 433
FT POLY-GLN.
FT CONFLICT 1 17
FT CONFLICT 117 118
FT CONFLICT 113 113
FT CONFLICT 1143 1380
FT N -> H (IN REF. 1).
FT NL -> IF (IN REF. 1).
FT N -> H (IN REF. 1).
FT RLIIPVYLFAKMRRLCLDLAHVIEIKLWDSNMNKALEEV
FT YDMSLREATEYALNMVDAMTFSFTYIKGKRRIFFPVFA
FT TTCMTAVLVISEYMKCVEDWARGYNANPNALIDFSRV
FT LWLKAEIRLRQNLIPKEDVLKSYTDFLRQDKDALDL
FT SALNEQAQRAMDPTDINETIQLIIVAAASLSKCLYLGV
FT LGDAPWPIILSFAGLQSLRATLYSVTKKRNTRI -> S
FT (IN REF. 1).
FT SEQUENCE 1380 AA; 155062 MW; FBE987EA5B639000 CRC64;

Query Match 63.2%; Score 55; DB 1; Length 1380;
Best Local Similarity 66.7%; Pred. No. 6.13e-01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1033 NGNEAYENC 1041
QY 63 NGDEAYDQC 71

RESULT 3
ID PYR1_SQUAC STANDARD; PRT; 2242 AA.
AC Q91437;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CAD PROTEIN [INCLUDES: GLUTAMINE-DEPENDENT CARBAMOYL-PHOSPHATE
DE SYNTHASE (EC 6.3.5.5); ASPARTATE CARBAMOYLTRANSFERASE (EC 2.1.
DE DIHYDROOROTASE (EC 3.5.2.3)].
GN CAD.
OS Squalus acanthias (Spiny dogfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondr
OC Elasmobranchii; Squaliformes; Squalidae; Squalus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS, AND SPLEEN;
RX MEDLINE; 95294021.
RA Hong J., Salo W.L., Anderson P.M.;
RT "Nucleotide sequence and tissue-specific expressi
RT multifunctional protein carbamoyl-phosphate syn
RT transcarbamoylase-dihydroorotase (CAD) mRNA in
RL J. Biol. Chem. 270:14130-14139(1995).
CC -!- FUNCTION: THIS PROTEIN IS A "FUSION" PROTEIN
CC ENZYMIC ACTIVITIES OF THE PYRIMIDINE
CC ATACASE AND THIOSE).
CC -!- CATALYTIC ACTIVITY: 2 ATP + GLUTAMINE +
CC ORTHOPHOSPHATE + GLUTAMATE + CARBAMOYL-PHOS
CC -!- CATALYTIC ACTIVITY: CARBAMOYL-PHOS
CC ORTHOPHOSPHATE + N-CARBAMOYLASPAR
CC -!- CATALYTIC ACTIVITY: DIHYDROOROT
CC L-ASPARTATE.
CC -!- COFACTOR: DHQASE ACTIVITY REQUIRING
CC -!- ENZYME REGULATION: ALLOSTERY
CC PHOSPHORYLATION. PRPP IS A
CC OF THE CPSASE REACTION (BMC.
CC -!- PATHWAY: FIRST TO THIRD IN THE
CC -!- SUBUNIT: HOMOHXAMER.
CC -!- SUBCELLULAR LOCATION:
CC -!- TISSUE SPECIFICITY: P/
```


RA Lozenan F.J., Litchfield D.W., Piening C., Takio K., Walsh K.A.,
 RA Krebs E.G.;
 RT "Isolation and characterization of human cDNA clones encoding the
 RT alpha and the alpha' subunits of casein kinase II.";
 RL Biochemistry 29:8436-8447(1990).
 CC -!- FUNCTION: CASEIN KINASES ARE OPERATIONALLY DEFINED BY THEIR
 CC PREFERENTIAL UTILIZATION OF ACIDIC PROTEINS SUCH AS CASEINS
 CC AS SUBSTRATES.
 CC -!- FUNCTION: THE ALPHA AND ALPHA' CHAINS CONTAIN THE CATALYTIC SITE.
 CC -!- SUBUNIT: TETRAMER COMPOSED OF AN ALPHA CHAIN, AN ALPHA' AND TWO
 CC BETA CHAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M55268; AA51548.1; -
 CC FIR; B35838; B35838.
 CC HSP; P28523; 1A60.
 CC MIM; 115442; -
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 CC PFAM; PF00069; pkinase; 1.
 CC Transferrase; Serine/threonine-protein kinase; ATP-binding.
 CC DOMAIN 40 325 PROTEIN_KINASE.
 CC NP_BIND 46 54 ATP (BY SIMILARITY).
 CC BINDING 69 69 ATP (BY SIMILARITY).
 CC ACT_SITE 157 157 BY SIMILARITY.
 CC SEQUENCE 350 AA; 41213 MW; 3ECB92F6BD3DD7F1 CRC64;
 CC
 CC Query Match 58.6%; Score 51; DB 1; Length 350;
 CC Best Local Similarity 55.6%; Pred. No. 4.39e+00;
 CC Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC Db 234 FHGDQNYDQ 242
 CC |::: |||
 CC Qy 62 FNGDEAYDQ 70
 CC
 CC RESULT 6
 CC ID KC22_CHICK STANDARD; PRT; 350 AA.
 CC AC P21869;
 CC DT 01-MAY-1991 (Rel. 18, Created)
 CC DT 01-MAY-1991 (Rel. 18, Last sequence update)
 CC DT 01-NOV-1995 (Rel. 32, Last annotation update)
 CC DE CASEIN KINASE II, ALPHA' CHAIN (CK II) (EC 2.7.1.37).
 CC OS Gallus gallus (Chicken).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Aves;
 CC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE; 91115855.
 CC RA Maridor G., Park W., Krek W., Nigg E.A.;
 CC RT "Casein kinase II. cDNA sequences, developmental expression, and
 CC RT tissue distribution of mRNAs for alpha, alpha', and beta subunits of
 CC RT the chicken enzyme.";
 CC RL J. Biol. Chem. 266:2362-2368(1991).
 CC
 CC -!- FUNCTION: CASEIN KINASES ARE OPERATIONALLY DEFINED BY THEIR
 CC PREFERENTIAL UTILIZATION OF ACIDIC PROTEINS SUCH AS CASEINS
 CC AS SUBSTRATES.
 CC -!- FUNCTION: THE ALPHA AND ALPHA' CHAINS CONTAIN THE CATALYTIC SITE.
 CC -!- SUBUNIT: TETRAMER COMPOSED OF AN ALPHA CHAIN, AN ALPHA' AND TWO
 CC BETA CHAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M59457; AAA48686.1; -
 CC FIR; B38611; B38611.
 CC HSP; P28523; 1A60.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 CC PFAM; PF00069; pkinase; 1.
 CC Transferrase; Serine/threonine-protein kinase; ATP-binding.
 CC DOMAIN 40 325 PROTEIN_KINASE.
 CC NP_BIND 46 54 ATP (BY SIMILARITY).
 CC BINDING 69 69 ATP (BY SIMILARITY).
 CC ACT_SITE 157 157 BY SIMILARITY.
 CC SEQUENCE 350 AA; 41246 MW; D0555CDE6CF78BFE CRC64;
 CC
 CC Query Match 58.6%; Score 51; DB 1; Length 350;
 CC Best Local Similarity 55.6%; Pred. No. 4.39e+00;
 CC Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC Db 234 FHGDQNYDQ 242
 CC |::: |||
 CC Qy 62 FNGDEAYDQ 70
 CC
 CC RESULT 7
 CC ID SCAG_HUMAN STANDARD; PRT; 649 AA.
 CC AC P51170; Q93023; Q93024; Q93025; Q93026; Q93027; P78437;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
 CC DE AMILORIDE-SENSITIVE SODIUM CHANNEL GAMMA-SUBUNIT (EPITHELIAL NA+
 CC DE CHANNEL GAMMA SUBUNIT) (GAMMA ENAC) (NONVOLTAGE-GATED SODIUM CHANNEL
 CC DE 1 GAMMA SUBUNIT) (SCNEG) (GAMMA NACH).
 CC GN SCNN1G.
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX TISSUE-LUNG;
 CC RX MEDLINE; 96039270.
 CC RA Voilley N., Bassilana F., Mignon C., Merscher S., Mattei M.-G.,
 CC Carle G.F., Lazdunski M., Barbry P.;
 CC RT "Cloning, chromosomal localization, and physical linkage of the beta
 CC RT and gamma subunits (SCNN1B and SCNN1G) of the human epithelial
 CC RT amiloride-sensitive sodium channel.";
 CC RL Genomics 28:560-565(1995).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RX TISSUE-KIDNEY;
 CC RX MEDLINE; 95282837.
 CC RA McDonald F.J., Snyder P.M., Price M.P., Welsh M.J.;
 CC RT "Cloning and expression of the beta- and gamma-subunits of the human
 CC RT epithelial sodium channel.";
 CC RL Am. J. Physiol. 268:C1157-C1163(1995).
 CC RN [3]
 CC RP SEQUENCE OF 1-43; 332-392; 433-463 AND 499-515 FROM N.A.
 CC RX MEDLINE; 96421599.
 CC RA Thomas C.P., Doggett N.A., Fisher R., Stokes J.B.;
 CC RT "Genomic organization and the 5' flanking region of the gamma subunit
 CC RT of the human amiloride-sensitive epithelial sodium channel.";
 CC RL J. Biol. Chem. 271:26062-26066(1996).
 CC CC -!- FUNCTION: SODIUM PERMEABLE NON-VOLTAGE-SENSITIVE ION CHANNEL
 CC OF THE LUMINAL SODIUM (AND WATER, WHICH FOLLOWS OSMOTICALLY)
 CC THROUGH THE APICAL MEMBRANE OF EPITHELIAL CELLS. CONTROLS THE
 CC REABSORPTION OF SODIUM IN KIDNEY, COLON, LUNG AND SWEAT GLANDS.
 CC ALSO PLAYS A ROLE IN TASTE PERCEPTION.
 CC -!- SUBUNIT: HETEROTRIMER OF AN ALPHA, BETA AND GAMMA SUBUNIT, A DELTA
 CC SUBUNIT CAN REPLACE THE ALPHA SUBUNIT.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- DISEASE: DEFECTS IN SCNNIG ARE ONE OF THE CAUSE OF LIDDLE SYNDROME
CC (PSEUDODYSTROPHISM), AN AUTOSOMAL DOMINANT FORM OF HYPERTENSION
CC ASSOCIATED WITH HYPOKALEMIC ALKALOSIS. THE DISEASE IS CAUSED BY
CC CONSTITUTIVE ACTIVATION OF THE RENAL EPITHELIAL SODIUM CHANNEL.
CC -1- SIMILARITY: BELONGS TO THE AMILORIDE-SENSITIVE SODIUM CHANNELS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X87160; CAA60633.1; -
CC EMBL; L36592; AAA75460.1; -
CC EMBL; U49336; AAC50737.1; -
CC EMBL; U53836; AAC50744.1; -
CC EMBL; U53845; AAC50753.1; -
CC EMBL; U53846; AAC50754.1; -
CC EMBL; U53847; AAC50755.1; -
CC EMBL; U53850; AAC50758.1; -
CC EMBL; U53852; AAC50760.1; -
CC MIM; 600761; -
CC PRINTS; PR01078; AMINACHANNEL.
CC PROSITE; PS01206; ASC; 1.
CC PFAM; PF00858; ASC; 1.
CC Ionic channel; Transmembrane; Ion transport; Glycoprotein.
CC DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 56 76 POTENTIAL.
CC DOMAIN 77 541 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 542 562 POTENTIAL.
CC DOMAIN 563 649 CYTOPLASMIC (POTENTIAL).
CC CARBOHYD 209 209 POTENTIAL.
CC CARBOHYD 248 248 POTENTIAL.
CC CARBOHYD 271 271 POTENTIAL.
CC CARBOHYD 291 291 POTENTIAL.
CC CARBOHYD 497 497 POTENTIAL.
CC CONFLICT 178 178 R -> W (IN REF. 2).
CC CONFLICT 339 339 F -> S (IN REF. 1).
CC CONFLICT 350 350 A -> T (IN REF. 1).
CC CONFLICT 369 369 Y -> S (IN REF. 1).
CC CONFLICT 375 375 D -> G (IN REF. 1).
CC CONFLICT 458 458 R -> S (IN REF. 2 AND 3).
CC CONFLICT 461 463 EWT -> DGH (IN REF. 3).
CC CONFLICT 502 502 P -> A (IN REF. 1).
CC CONFLICT 614 614 A -> S (IN REF. 2).
CC SEQUENCE 649 AA; 74365 MW; 02DE31178A74D2FF CRC64;
Query Match 58.6%; Score 51; DB 1; Length 649;
Best Local Similarity 50.0%; Pred. No. 4.39e+00;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
Db 363 FKLSEPSQCTE 374
QY 62 FNGDEAYDQCTN 73
RESULT 8 STANDARD; PRT; 650 AA.
AC P37091;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE AMILORIDE-SENSITIVE SODIUM CHANNEL GAMMA-SUBUNIT (EPITHELIAL NA+
DE CHANNEL GAMMA SUBUNIT) (GAMMA ENAC) (NONVOLTAGE-GATED SODIUM CHANNEL
DE 1 GAMMA SUBUNIT) (SCNEG) (GAMMA NACH).
GN SCNNIG.
OS Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
CC

CC RN SEQUENCE FROM N.A.
CC RP STRAIN-SPRAGUE-DAWLEY; TISSUE-DISTAL COLON EPITHELIUM;
CC RX MEDLINE; 94150624.
CC RA Canessa C.M., Schild L., Buell G., Thorens B., Gautschi I.,
CC RA Horisberger J.D., Rossier B.C.;
CC RT "Amiloride-sensitive epithelial Na+ channel is made of three
CC RT homologous subunits.";
CC RL Nature 367:463-467(1994).
CC [2]
CC SEQUENCE FROM N.A.
CC RP STRAIN-WISTAR; TISSUE-DISTAL COLON;
CC RX MEDLINE; 94245676.
CC RA Lingueglia R., Renard S., Waldmann R., Voilley N., Champigny G.,
CC RA Plass H., Lazdunski R., Barbry P.;
CC RT "Different homologous subunits of the amiloride-sensitive Na+ channel
CC RT are differentially regulated by aldosterone.";
CC RL J. Biol. Chem. 269:13736-13739(1994).
CC [3]
CC SEQUENCE FROM N.A.
CC RP STRAIN-SPRAGUE-DAWLEY, ABD SHRSHPD; TISSUE-DISTAL COLON, AND KIDNEY;
CC RX MEDLINE; 97191134.
CC RA Kreutz R., Struk B., Rubattu S., Hubner N., Szpirer J.,
CC RA Szpirer C., Ganten D., Lindpaintner K.;
CC RT "Role of the alpha-, beta-, and gamma-subunits of epithelial sodium
CC RT channel in a model of polygenic hypertension.";
CC RL Hypertension 29:131-136(1997).
CC -1- FUNCTION: SODIUM PERMEABLE NON-VOLTAGE-SENSITIVE ION CHANNEL
CC INHIBITED BY THE DIURETIC AMILORIDE. MEDIATES THE ELECTRODIFFUSION
CC OF THE LUMINAL SODIUM (AND WATER, WHICH FOLLOWS OSMOTICALLY)
CC THROUGH THE APICAL MEMBRANE OF EPITHELIAL CELLS. CONTROLS THE
CC REABSORPTION OF SODIUM IN KIDNEY, COLON, LUNG AND SWEAT GLANDS.
CC ALSO PLAYS A ROLE IN TASTE PERCEPTION.
CC -1- SUBUNIT: HETEROTRIMER OF AN ALPHA, BETA AND GAMMA SUBUNIT, A DELTA
CC SUBUNIT CAN REPLACE THE ALPHA SUBUNIT.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE AMILORIDE-SENSITIVE SODIUM CHANNELS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X77933; CAA54505.1; -
CC EMBL; X78034; CAA54964.1; -
CC EMBL; U37539; AAB58459.1; -
CC EMBL; U37540; AAB58460.1; -
CC PIR; S41160; S41160.
CC PRINTS; PR01078; AMINACHANNEL.
CC PROSITE; PS01206; ASC; 1.
CC PFAM; PF00858; ASC; 1.
CC Ionic channel; Transmembrane; Ion transport; Glycoprotein.
CC DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 56 76 POTENTIAL.
CC DOMAIN 77 542 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 543 563 POTENTIAL.
CC DOMAIN 564 650 CYTOPLASMIC (POTENTIAL).
CC CARBOHYD 210 210 POTENTIAL.
CC CARBOHYD 249 249 POTENTIAL.
CC CARBOHYD 272 272 POTENTIAL.
CC CARBOHYD 292 292 POTENTIAL.
CC CARBOHYD 498 498 POTENTIAL.
CC CONFLICT 53 53 R -> P (IN REF. 1).
CC CONFLICT 573 573 W -> C (IN REF. 1).
CC SEQUENCE 650 AA; 74066 MW; 701F9B28B3250D8F CRC64;
Query Match 58.6%; Score 51; DB 1; Length 650;
Best Local Similarity 50.0%; Pred. No. 4.39e+00;


```

AC P08181;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE CASEIN KINASE II, ALPHA CHAIN (CK II) (EC 2.7.1.37).
GN CASK-II-A.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88063475.
RA Saxena A., Padmanabha R., Glover C.V.C.;
RT "Isolation and sequencing of cDNA clones encoding alpha and beta
RL subunits of Drosophila melanogaster casein kinase II.";
RL Mol. Cell. Biol. 7:3409-3417(1987).
RN [2]
RP SEQUENCE OF 1-41.
RX MEDLINE: 87109340.
RA Padmanabha R., Glover C.V.C.;
RT "Casein kinase II of yeast contains two distinct alpha polypeptides
RT and an unusually large beta subunit.";
RL J. Biol. Chem. 262:1829-1835(1987).
CC -!- FUNCTION: CASEIN KINASES ARE OPERATIONALLY DEFINED BY THEIR
CC PREFERENTIAL UTILIZATION OF ACIDIC PROTEINS SUCH AS CASEINS
CC AS SUBSTRATES.
CC -!- FUNCTION: THE ALPHA CHAIN CONTAINS THE CATALYTIC SITE.
CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M16534; AAA28429.1; -
DR PIR: A26688; A26688.
DR HSSP: P28523; 1A60.
DR FLYBASE: FBgn0000258; CkII-alpha.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM: PF00069; pkinase; 1.
DR TRANSFERASE; Serine/threonine-protein kinase; ATP-binding.
FT INIT_MET 0
FT DOMAIN 36 321 PROTEIN KINASE.
FT NP_BIND 42 50 ATP (BY SIMILARITY).
FT BINDING 65 65 ATP (BY SIMILARITY).
FT ACT_SITE 153 153 BY SIMILARITY.
SQ SEQUENCE 335 AA; 39828 MW; 46A478E3076D85A CRC64;

Query Match 57.5%; Score 50; DB 1; Length 335;
Best Local Similarity 55.6%; Pred. No. 7.05e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 230 FHGHNDYDQ 238
QY 62 FNGDEAYDQ 70

RESULT 11
ID KC22_XENLA STANDARD; PRT; 350 AA.
AC P28020;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE CASEIN KINASE II, ALPHA' CHAIN (CK II) (EC 2.7.1.37).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;

```

```

OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-Ovary.
RX MEDLINE: 92183811.
RA Jedlicki A., Hiorichs M.V., Allende C., Allende J.E.;
RT "The cDNAs coding for the alpha- and beta-subunits of Xenopus laevis
RL casein kinase II.";
RL FEBS Lett. 297:280-284(1992).
CC -!- FUNCTION: CASEIN KINASES ARE OPERATIONALLY DEFINED BY THEIR
CC PREFERENTIAL UTILIZATION OF ACIDIC PROTEINS SUCH AS CASEINS
CC AS SUBSTRATES.
CC -!- FUNCTION: THE ALPHA AND ALPHA' CHAINS CONTAIN THE CATALYTIC SITE.
CC -!- SUBUNIT: TETRAMER COMPOSED OF AN ALPHA CHAIN, AN ALPHA' AND TWO
CC BETA CHAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62375; CAA44238.1; -
DR PIR: S18897; S18897.
DR PIR: S20404; S20404.
DR HSSP: P28523; 1A60.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM: PF00069; pkinase; 1.
DR TRANSFERASE; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 39 324 # PROTEIN KINASE.
FT NP_BIND 45 53 ATP (BY SIMILARITY).
FT BINDING 68 68 ATP (BY SIMILARITY).
FT ACT_SITE 156 156 BY SIMILARITY.
SQ SEQUENCE 350 AA; 41454 MW; B317B58CFE5DAA67 CRC64;

Query Match 57.5%; Score 50; DB 1; Length 350;
Best Local Similarity 55.6%; Pred. No. 7.05e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 233 FHGHNDYDQ 241
QY 62 FNGDEAYDQ 70

RESULT 12
ID GBA1_DICDI STANDARD; PRT; 356 AA.
AC P16894;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE GUANINE NUCLEOTIDE-BINDING PROTEIN ALPHA-1 SUBUNIT (G ALPHA 1).
GN GPAA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Dictyosteliida; Dictyostelium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89296910.
RA Pupillo M., Kumagai A., Pitt G.S., Firtel R.A., Devreotes P.N.;
RT "Multiple alpha subunits of guanine nucleotide-binding proteins in
RT Dictyostelium";
RT Dictyostelium.
RL Proc. Natl. Acad. Sci. U.S.A. 86:4892-4896(1989).
RN [2]
RP SEQUENCE OF 5-35; FROM N.A.
RX MEDLINE: 89195237.
RA Kumagai A., Pupillo M., Gundersen R., Miske-Lye R., Devreotes P.N.;
RA Firtel R.A.;
RT "Regulation and function of G alpha protein subunits in
RT Dictyostelium.";
RL Cell 57:265-275(1989).

```


Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 233 FHGHNDYDQ 241
|:|:|
QY 62 FNGDEAYDQ 70

RESULT 15

ID KC21_RABIT STANDARD; PRT; 391 AA.
AC P33674;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CASEIN KINASE II, ALPHA CHAIN (CK II) (EC 2.7.1.37).
GN CSNK2A1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93185937.
RA Gupta S.K., Singh J.P.;
RT "PCR cloning and sequence of two cDNAs encoding the alpha and beta
RT subunits of rabbit casein kinase-II.";
RL Gene 124:287-290(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NEW ZEALAND WHITE;
RA Gupta S.K., Rothfuss K.J., Singh J.P.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CASEIN KINASES ARE OPERATIONALLY DEFINED BY THEIR
CC PREFERENTIAL UTILIZATION OF ACIDIC PROTEINS SUCH AS CASEINS
CC AS SUBSTRATES.
CC -!- FUNCTION: THE ALPHA AND ALPHA' CHAINS CONTAIN THE CATALYTIC SITE.
CC -!- SUBUNIT: TETRAMER COMPOSED OF AN ALPHA CHAIN, AN ALPHA' AND TWO
CC BETA CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; S56527; AAB25554.1; -;
EMBL; M98451; AAB1891.1; -;
PIR; JN0555; JN0555.
HSP; P28523; 1A6O.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinase; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 39
FT NP_BIND 45 53 ATP (BY SIMILARITY).
FT BINDING 68 68 ATP (BY SIMILARITY).
FT ACT_SITE 156 156 BY SIMILARITY.
SQ SEQUENCE 391 AA; 45115 MW; 5212943C906905D4 CRC64;

Query Match 57.5%; Score 50; DB 1; Length 391;
Best Local Similarity 55.6%; Pred. No. 7.05e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 233 FHGHNDYDQ 241
|:|:|
QY 62 FNGDEAYDQ 70

Search completed: Wed May 10 12:29:16 2000
Job time : 100 secs.

M P S R E L A

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 12:29:34 2000; MasPar time 227.48 Seconds
3.657 Million cell updates/sec

Tabular output not generated.

Title: >US-09-376-430-2
Description: (62-73) from US09376430A.pep (11 of 25)
Perfect Score: 87
Sequence: 1 FNGDEAYDQCTN 12

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 23.577; Variance 28.405; scale 0.830

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Description	Pred. No.
1	55	63.2	85 10 Q39943	HOMEODOMAIN CONTAINING
2	54	62.1	679 13 Q98856	GELATINASE-B.
3	53	60.9	504 5 Q94470	GTP-BINDING PROTEIN AL
4	52	59.8	103 5 Q62121	D1025.4 PROTEIN.
5	52	59.8	288 10 Q9X119	T10024.25
6	51	58.6	296 13 Q91397	CASEIN KINASE 2 ALPHA
7	51	58.6	348 13 Q90457	PROTEIN KINASE CK2 ALP
8	51	58.6	350 11 Q54833	CASEIN KINASE II, ALPH
9	51	58.6	359 5 Q62433	Y41E3.3 PROTEIN.
10	51	58.6	441 14 Q91257	3D POLYMERASE (FRAGMEN
11	51	58.6	655 11 Q9WU39	EPITHELIAL SODIUM CHAN
12	51	58.6	713 3 Q42947	HYPOTHETICAL 80.1 KD P
13	50	57.5	213 2 Q25918	RESPONSE REGULATOR.
14	50	57.5	353 5 Q76484	CASEIN KINASE II ALPHA
15	50	57.5	391 11 Q61177	CASEIN KINASE II, ALPH
16	50	57.5	391 4 Q14013	CASEIN KINASE II ALPHA
17	50	57.5	460 14 Q89222	ORF1L (FRAGMENT).
18	50	57.5	680 2 Q66812	OUTER MEMBRANE PROTEIN
19	49	56.3	212 5 Q17223	ND-SD MUTANT FIBROIN L
20	49	56.3	276 5 Q17221	ND-S MUTANT FIBROIN LI

21 49 56.3 276 5 Q17222
22 49 56.3 277 2 Q53640
23 49 56.3 343 2 P72464
24 49 56.3 379 3 Q59934
25 49 56.3 404 5 Q01428
26 49 56.3 554 5 Q18253
27 49 56.3 555 5 Q18351
28 49 56.3 1259 14 Q9YKC5
29 49 56.3 1336 3 Q42803
30 49 56.3 1356 3 Q42731
31 49 56.3 1502 3 Q60081
32 48 55.2 172 2 Q25810
33 48 55.2 178 2 P71406
34 48 55.2 316 3 P78755
35 48 55.2 317 3 Q42970
36 48 55.2 348 2 Q66408
37 48 55.2 394 2 Q51724
38 48 55.2 437 10 Q81041
39 48 55.2 470 5 P91086
40 48 55.2 596 14 Q9YMI0
41 48 55.2 643 14 Q90211
42 48 55.2 643 14 Q83249
43 48 55.2 1707 14 Q67704
44 48 55.2 2109 13 P79787
45 48 55.2 2664 5 Q26033

ALIGNMENTS

RESULT 1
ID Q39943 PRELIMINARY: PRT; 85 AA.
AC Q39943;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE HOMEODOMAIN CONTAINING PROTEIN (HAHB-3) (FRAGMENT).
GN HAHB-3.
OS Helianthus annuus (Common sunflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids II; Asterales; Asteraceae;
OC Asteroideae; Heliantheae; Helianthus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=STEM;
RX MEDLINE; 95148747.
RA CHAN R.L., GONZALEZ D.H.;
RT "A cDNA encoding an HD-zip protein from sunflower.";
RL Plant Physiol. 106:1687-1688(1994).
DR EMBL; L22848; AAA63767.1; -;
DR MENDEL; 8451; Helan;1345;8451.
KW Homeobox; DNA-binding; Nuclear protein.
FT NON_TER 1 85
FT NON_TER 85 85
SQ SEQUENCE 85 AA; 10213 MW; 9FCAA98 CRC32;
Query Match 63.2%; Score 55; DB 10; Length 85;
Best Local Similarity 50.0%; Pred.No. 1.07e+00;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 26 YNSDEYDENC 35
QY 62 FNGDEAYDQC 71
RESULT 2
ID Q98856 PRELIMINARY: PRT; 679 AA.
AC Q98856;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE GELATINASE-B.
OS Cynops pyrrhogaster (Japanese common newt).

[illegible]


```

DE CASEIN KINASE 2 ALPHA SUBUNIT (FRAGMENT).
GN CK2A1.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprininae; Rasbora; Danio.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95253232.
RA DANIORTI J.L., ALLENDE M.L., WEINBERG E.S., ALLENDE J.E.;
RT "Cloning and expression of genes coding for protein kinase CK2 alpha
RT and beta subunits in zebrafish (Danio rerio).";
RL Cell. Mol. Biol. Res. 40:431-439(1994).
DR EMBL; S76875; AAB34248.1; -.
DR HSSP; P28523; 1A60.
DR ZFIN; ZDB-GENE-990415-27; ck2a1.
DR PFAM; PF00069; pkinase; 1.
FT NON_TER 1
SQ SEQUENCE 296 AA; 34662 MW; 678F8611 CRC32;

Query Match 58.6%; Score 51; DB 13; Length 296;
Best Local Similarity 55.6%; Pred. No. 7.63e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 171 FHGQNDYDQ 179
QY 62 FNGDEAYDQ 70

RESULT 7
ID Q90457 PRELIMINARY; PRT; 348 AA.
AC Q90457;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE PROTEIN KINASE CK2 ALPHA.
GN CK2A2 OR CK2 ALPHA.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprininae; Rasbora; Danio.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97054619.
RA ANTONELLI M., DANIORTI J.L., ROJO D., ALLENDE C.C., ALLENDE J.E.;
RT "Cloning, expression and properties of the alpha' subunit of casein
RT kinase 2 from zebrafish (Danio rerio).";
RL Eur. J. Biochem. 241:272-279(1996).
DR EMBL; X99964; CAA68229.1; -.
DR HSSP; P28523; 1A60.
DR ZFIN; ZDB-GENE-990415-28; ck2a2.
DR PFAM; PF00069; pkinase; 1.
KW Kinase.
SQ SEQUENCE 348 AA; 40877 MW; D3AA96C5 CRC32;

Query Match 58.6%; Score 51; DB 13; Length 348;
Best Local Similarity 55.6%; Pred. No. 7.63e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 233 FHGQNDYDQ 241
QY 62 FNGDEAYDQ 70

RESULT 8
ID O54833 PRELIMINARY; PRT; 350 AA.
AC O54833;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE CASEIN KINASE II, ALPHA 2, POLYPEPTIDE
DE (CASEIN KINASE II, ALPHA PRIME SUBUNIT).
GN CSNK2A2 OR CSKN2A2.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98163749.
RA XU X., RICH E.S. JR., SELDIN D.C.;
RT "Murine protein kinase CK2 alpha': cDNA and genomic cloning and
RT chromosomal mapping.";
RL Genomics 48:79-86(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6;
RX STRAIN=C57;
RX MEDLINE; 98362010.
RA ORLANDINI M., SEMPLICI F., FERRUZZI R., MEGGIO F., PINNA L.A.,
RA OLIVIERO S.;
RT "Protein kinase CK2alpha' is induced by serum as a delayed early gene
RT and cooperates with Ha-ras in fibroblast transformation.";
RL J. Biol. Chem. 273:21291-21297(1998).
DR EMBL; AF012251; AAC53552.1; -.
DR EMBL; AJ001420; CAA04753.1; -.
DR HSSP; P28523; 1A60.
DR MGD; MGI:88547; Csnk2a2.
DR PFAM; PF00069; pkinase; 1.
SQ SEQUENCE 350 AA; 41215 MW; B7965739 CRC32;

Query Match 58.6%; Score 51; DB 11; Length 350;
Best Local Similarity 55.6%; Pred. No. 7.63e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 234 FHGQNDYDQ 242
QY 62 FNGDEAYDQ 70

RESULT 9
ID O62433 PRELIMINARY; PRT; 359 AA.
AC O62433;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-JAN-1999 (TRENBLrel. 09, Last annotation update)
DE Y41E3.3 PROTEIN.
GN Y41E3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderidae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., FULTON L.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z95559; CAB09000.1; -.
SQ SEQUENCE 359 AA; 40395 MW; 56720894 CRC32;

Query Match 58.6%; Score 51; DB 5; Length 359;
Best Local Similarity 40.0%; Pred. No. 7.63e+00;
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

```

Db 7 FDNSTYQOC 16
QY 62 FNGDEAYDQC 71

RESULT 10
ID O91257; PRELIMINARY; PRT; 441 AA.
AC O91257;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE 3D POLYMERASE (FRAGMENT).
OS Porcine enterovirus 8.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V13;
RA PENG J.H., LIN F., KITCHING R.P., KNOWLES N.J.;
RT "Nucleotide and deduced amino acid sequence of the 3' end of the
RT porcine enterovirus type 8 genome.";
RT Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ001391; CAA04721.1;
DR PFAM: PF00680; RNA_dep_RNA_poi; 1.
ET NON_TER
SQ SEQUENCE 441 AA; 49746 MW; 67A3F07C CRC32;

Query Match 58.6%; Score 51; DB 14; Length 441;
Best Local Similarity 50.0%; Pred.No. 7.63e+00;
Matches 6; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

Db 404 NGEVYVECTK 415
QY 63 FNGDEAYDQ-CTN 73

RESULT 11
ID Q9WU39 PRELIMINARY; PRT; 655 AA.
AC Q9WU39;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE EPITHELIAL SODIUM CHANNEL GAMMA SUBUNIT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6;
RA AHN Y.J., BROOKER D.B., HARTE B.J., KOSARI F., LI J., MACKLER S.A.,
RA KLEYMAN T.R.;
RT "Cloning and functional expression of the mouse epithelial sodium
RT channel.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE AMILORIDE-SENSITIVE SODIUM CHANNELS
CC FAMILY.
DR EMBL: AF112187; AAD21246.1;
DR PROSITE; PS01206; ASC; 1.
DR Ionic channel; Transmembrane; Ion transport.
KW SEQUENCE 655 AA; 74635 MW; FB87DBA6 CRC32;

Query Match 58.6%; Score 51; DB 11; Length 655;
Best Local Similarity 50.0%; Pred.No. 7.63e+00;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 369 FKLSEPVSOCTE 380
QY 62 FNGDEAYDQCTN 73

RESULT 12

```

```

ID O42947 PRELIMINARY; PRT; 713 AA.
AC O42947;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JUN-1998 (TREMBlrel. 06, Last annotation update)
DE HYPOTHETICAL 80.1 KD PROTEIN.
GN SPBC16H5.12C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-;
RA WOOD V., RAJANDREAM M.A., BARRELL B.G., SKELTON J., CHURCHER C.M.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL022104; CAA17910.1;
KW Hypothetical protein.
SQ SEQUENCE 713 AA; 80067 MW; 826B7F44 CRC32;

Query Match 58.6%; Score 51; DB 3; Length 713;
Best Local Similarity 30.0%; Pred.No. 7.63e+00;
Matches 3; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 561 YSGEQRYERC 570
QY 62 FNGDEAYDQC 71

RESULT 13
ID O25918 PRELIMINARY; PRT; 213 AA.
AC O25918;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE RESPONSE REGULATOR.
GN HPI365.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-26695;
RX MEDLINE; 97394467.
RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA VENTER J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
DR EMBL: AE000636; AAD08405.1;
DR HSSP; P06628; 2FSP.
DR TIGR; HP1365;
DR PFAM; PF00072; response_reg; 1.
DR PFAM; PF00486; trans_reg_C; 1.
KW Hypothetical protein.
SQ SEQUENCE 213 AA; 24871 MW; CC138F1E CRC32;

Query Match 57.5%; Score 50; DB 2; Length 213;
Best Local Similarity 66.7%; Pred.No. 1.23e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 33 FNGKEAYER 41
QY 62 FNGDEAYDQ 70

```

```

RESULT 14
ID O76484 PRELIMINARY; PRT; 353 AA.
AC O76484;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE CASEIN KINASE II ALPHA SUBUNIT.
OS Spodoptera frugiperda (Fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Noctuoidea; Noctuidae; Amphipyrinae; Spodoptera.
RN [1]
SEQUENCE FROM N.A.
RA EBEL W., STEPLEWSKI A., ROBERTSON N.M., ALNEMRI E.S., LITWACK G.;
RT "Identification, cloning and expression of Spodoptera frugiperda
RT casein Kinase II."
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF071210; AAC24041.1; -.
DR HSSP; P28523; IA60.
DR PFAM; PF00069; pkinase; 1.
SQ SEQUENCE 353 AA; 42025 MW; AC8C2E11 CRC32;

```

```

Query Match 57.5%; Score 50; DB 5; Length 353;
Best Local Similarity 55.6%; Pred.No. 1.23e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 233 FHGHNDYDQ 241
|:|:|
QY 62 FNGDEAYDQ 70

```

```

RESULT 15
ID Q61177 PRELIMINARY; PRT; 391 AA.
AC Q61177;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE CASEIN KINASE II, ALPHA 1, RELATED SEQUENCE 4
DE (CASEIN KINASE II ALPHA SUBUNIT).
GN CSNK2A1-RS4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
SEQUENCE FROM N.A.
RP STRAIN-B10.A/AKR;
RA CHEN L., HARDWICK J.P., SITKOVSKY M.V.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U51866; AAA96795.1; -.
DR HSSP; P28523; IA60.
DR MGD; MGI:894650; Csnk2al-rs4.
DR PFAM; PF00069; pkinase; 1.
SQ SEQUENCE 391 AA; 45179 MW; 888B30D7 CRC32;

```

```

Query Match 57.5%; Score 50; DB 11; Length 391;
Best Local Similarity 55.6%; Pred.No. 1.23e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 233 FHGHNDYDQ 241
|:|:|
QY 62 FNGDEAYDQ 70

```

Search completed: Wed May 10 12:33:33 2000
Job time : 239 secs.

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 12:42:29 2000; MasPar time 2.80 Seconds
Tabular output not generated. 67.656 Million cell updates/sec

Title: >US-09-376-430-2
Description: (78-85) from US09376430A.pep (12 of 25)
Perfect Score: 59
Sequence: 1 EGHVSGCL 8

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:geneseq

Statistics: Mean 14.702; Variance 41.023; scale 0.358

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query Match	Length	ID	Description	Pred. No.	
1	44	74.6	1239	1 W56131	Drosophila melanogaster	1.28e+02	
2	43	72.9	902	1 P93110	Rat androgen receptor	1.67e+02	
3	43	72.9	902	1 R12224	Rat androgen receptor	1.67e+02	
4	43	72.9	902	1 P31006	Rat androgen receptor	1.67e+02	
5	43	72.9	1801	1 W50895	Rat laminin B2 chain	1.67e+02	
6	42	71.2	196	1 R05285	Amino acid sequence of	2.18e+02	
7	42	71.2	197	1 R39228	Recombinant human alph	2.18e+02	
8	42	71.2	340	1 W5731	H. pylori ORF 14cp1190	2.18e+02	
9	42	71.2	387	1 W98493	H. pylori GHPO 1072 pr	2.18e+02	
10	42	71.2	389	1 R39222	Recombinant human alph	2.18e+02	
11	42	71.2	389	1 W12718	PhC gene product	2.18e+02	
12	42	71.2	590	1 W10123	Alpha-foetoprotein	2.18e+02	
13	42	71.2	609	1 R39227	Human alpha-foetoprote	2.18e+02	
14	41	69.5	28	1 R31385	ME#1 library derived p	2.84e+02	
15	41	69.5	328	1 W19761	Pyruvate:ferredoxin ox	2.84e+02	
16	41	69.5	440	1 W24789	Human lecithin-cholest	2.84e+02	
17	41	69.5	440	1 P70134	Natural recombinant hu	2.84e+02	
18	40	67.8	28	1 R31402	ME#1 library derived p	3.68e+02	
19	40	67.8	153	1 W35692	Human endogenous retro	3.68e+02	
20	40	67.8	153	1 W37745	Human endogenous retro	3.68e+02	
21	40	67.8	181	1 W37747	Human endogenous retro	3.68e+02	
22	40	67.8	181	1 W95694	Human endogenous retro	3.68e+02	
23	40	67.8	436	1 W37350	Psba (WbpA) protein in	3.68e+02	

24	40	67.8	1	W56131	standard; Protein; 1239 AA.
25	40	67.8	1	W56131	
26	40	67.8	1	W56131	
27	40	67.8	1	W56131	
28	39	66.1	74	1 W719	Branched chain keto ac
29	39	66.1	192	1 R6964	Human endogenous retro
30	39	66.1	192	1 R8951	Human endogenous retro
31	39	66.1	261	1 R22960	Perigillus nidulans m
32	39	66.1	261	1 R21517	Coccyte nucleolar m
33	39	66.1	340	1 W36598	C virus enve
34	39	66.1	377	1 W21965	C virus enve
35	39	66.1	425	1 R13792	C virus enve
36	39	66.1	580	1 W25670	C virus enve
37	39	66.1	580	1 W46758	C virus enve
38	39	66.1	796	1 W71080	Asome isol
39	39	66.1	796	1 W36024	MSK1 protease
40	39	66.1	861	1 W22232	MSK1 beta-3
41	39	66.1	861	1 W14138	Human beta-3
42	39	66.1	904	1 R05619	Human beta-3
43	39	66.1	1609	1 W50898	BIV enve
44	38	64.4	2458	1 R07640	Human l
45	38	64.4	2813	1 W54347	Deduced

ALIGNMENTS

RESULT 1
ID W56131 standard; Protein; 1239 AA.
AC W56131
DT 17-AUG-1998 (first entry)
DE Drosophila melanogaster KUZ protein.
KW kuzbanian; kuz; neurogenic; KUZ protein; neuronal partition
KW development; NOTCH protein processing; regulation; cell func
KW signal transduction pathways; screening; receptor binding;
KW metalloprotease.
OS Drosophila melanogaster. #
PN WO9808933-Al.
PD 05-MAR-1998.
PF 27-AUG-1997; U15099.
PR 23-JUL-1997; US-019390.
PR 29-AUG-1996; US-019390.
PA (RESC) UNIV CALIFORNIA.
PA (UYIA) UNIV YALE.
PI Pan D, Rooke J, Rubin GM, Xu T, Yavari R;
DR WPI; 98-179428/16.
DR N-PSDB; V22647.
PT New KUZ polypeptides, members of the ADAM family of metalloprotease
PT - useful in neural partitioning and development
PS Claim 1; Pages 31-34; 58pp; English.
CC The sequence is that of encoded by the kuzbanian gene, a
CC neurogenic gene. The KUZ protein is involved in neuronal
CC partitioning and development. It is also involved in processing
CC of the NOTCH protein by cleaving the C-terminal portion. The
CC KUZ protein can be used to screen for compounds that alter
CC binding of KUZ to its receptor or the cleavage of the NOTCH
CC protein, hence acting to regulate NOTCH signal transduction
CC pathways and regulate cell functions.
SQ Sequence 1239 AA;

Query Match 74.6%; Score 44; DB 1; Length 1239;
Best Local Similarity 71.4%; Pred. No. 1.28e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 273 EGHVAGC 279
QY 78 EGHVAGC 84

RESULT 2
ID P93110 standard; protein; 902 AA.
AC P93110
DT 19-MAR-1990 (first entry)
DE Rat androgen receptor.

KW Rat androgen receptor; monoclonal antibody; polyclonal antibody; cancer.
 OS Rattus rattus.
 PN WO8909791-A.
 PD 19-OCT-1989.
 PF 13-APR-1989; U01548.
 PR 14-APR-1988; US-182646.
 PA (UYNC-) University of North Carolina.
 PI French FS, Wilson EM, Joseph DR, Lubahn DB;
 DR WPI: 89-324206/44.
 DR N-PSDB: N91773.
 PT DNA encoding androgen receptor protein - useful for transforming
 PT eukaryotic hosts for protein expression and subsequent antibody prodn.
 PS Disclosure; Fig. 5; 41pp; English.
 CC Androgen receptor protein (AR) is used to produce mono- or poly-clonal
 CC antibodies. These are used for the detection and quantification of AR in
 CC the presence of endogenous androgen, as androgen will not interfere with
 CC binding. They may be used in assays to determine and quantify cellular
 CC distribution of AR in tumour tissue, and are esp. useful for evaluating
 CC prostate cancers to determine responsiveness to androgen withdrawal
 CC therapy.
 SQ Sequence 902 AA;

Query Match 72.9%; Score 43; DB 1; Length 902;
 Best Local Similarity 77.8%; Pred. No. 1.67e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 103 EGHPEGCL 111
 ||| ||||
 QY 78 EGH-T-SGCL 85

RESULT 3

ID R12224 standard; Protein; 902 AA.
 AC R12224;
 DT 20-AUG-1991 (first entry)
 DE Rat androgen receptor.
 KW RAR; DNA-binding protein; steroid hormone.
 OS Rattus rattus.
 FH Key Location/Qualifiers
 FT domain 540..610
 FT /label= "DNA-binding domain"
 FT /note= "cysteine-rich"

WO9107423-A.

PN 30-MAY-1991.
 PD 19-OCT-1990; U06015.
 PF 17-NOV-1989; US-438775.
 PR (ARCH-) ARCH DEV CORP.
 PA Liao S, Chang C;
 PI WPI: 91-178048/24.
 DR N-PSDB: Q12002.
 PT Androgen receptor and TR2 DNA binding proteins - DNA sequences
 PT and antibodies for detection and quantification methods
 PS Claim 25; Fig 3; 79pp; English.
 CC This sequence was deduced from a cDNA clone isolated by screening
 CC a rat ventral prostate lambda gt10 library in E.coli Y1090. The
 CC sequence is very similar to that of human AR and in
 CC the DNA-binding domain it is identical to that of hAR DNA-binding
 CC domain. Homology comparisons with other known steroid receptors
 CC indicate that RAR is more closely related to glucocorticoid,
 CC mineralo-corticoid and progesterone receptors than to v-erb-A or to
 CC receptors for oestrogen, vitamin D and thyroid hormones.
 SQ Sequence 902 AA;

Query Match 72.9%; Score 43; DB 1; Length 902;
 Best Local Similarity 77.8%; Pred. No. 1.67e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 103 EGHPEGCL 111
 ||| ||||
 QY 78 EGH-T-SGCL 85

RESULT 4

ID P91006 standard; protein; 902 BP.
 AC P91006;
 DT 28-FEB-1990 (first entry)
 DE Rat androgen receptor DNA clone.
 KW Androgen receptor; TR2 polypeptide.
 OS Rat.
 FH Key Location/Qualifiers
 FT region 1..902
 FT /tag= a
 FT /product=98 kD polypeptide
 FT 170..902
 FT /tag= b
 FT /product=79 kD polypeptide

WO8909223-A.

PN 05-OCT-1989.
 PD 24-MAR-1989; U01238.
 PF 30-MAR-1988; US-176107.
 PR (ARCH-) Arch Development Corp.
 PA Liao S, Chang C;
 PI WPI: 89-309501/42.
 DR N-PSDB: N91578.
 PT New DNA encoding new androgen receptor and TR2 polypeptide(s) - able
 PT to bind DNA, and derived antibodies, useful for receptor assay and
 PT purification.
 PS Claim 8; fig. 3; 60pp; English.
 CC The protein is used to raise antibodies for receptor assays and for
 CC affinity purification.
 CC The 98 kD product starts at the first Met codon; the 79 kD product
 CC starts from the second.
 SQ Sequence 902 AA;

Query Match 72.9%; Score 43; DB 1; Length 902;
 Best Local Similarity 77.8%; Pred. No. 1.67e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 103 EGHPEGCL 111
 ||| ||||
 QY 78 EGH-T-SGCL 85

RESULT 5

ID W50895 standard; Protein; 1801 AA.

AC W50895;
 DT 07-DEC-1998 (first entry)
 DE Rat laminin B2 chain.
 KW Laminin; rat; beta-amyloid; amyloidosis; Alzheimer's disease;
 KW Down's syndrome; hereditary cerebral haemorrhage; inflammatory
 KW malignancy; Familial Mediterranean Fever; multiple myeloma;
 KW type II diabetes; prion disease; Creutzfeldt-Jacob disease;
 KW Gerstmann-Strausler syndrome; kuru; scrapie; haemodial;
 KW carpal tunnel syndrome; senile cardiac amyloid polynour;
 KW Familial Amyloidotic Polyneuropathy; thyroid carcinoma
 KW therapy.

OS Rattus sp.
 PN WO9815179-A1.
 PD 16-APR-1998.
 PF 08-OCT-1997; U18145.
 PR 08-OCT-1996; US-027981.
 PA (UNIW) UNIV WASHINGTON.
 PI Castillo G, Snow AD;
 DR WPI: 98-240534/21.
 PT Use of laminin and fragments - for development as potent
 PT the diagnosis and treatment of amyloidosis. The
 PT disease or CJD
 PS Claim 15; page 94-97; 132pp; English; accumulates laminin
 CC This is the amino acid sequence of other mammalian laminin
 CC primary object of the invention and other mammalian (merosin)
 CC protein fragments and/or laminin include in A2 chain A1 chain
 CC inhibitors of amyloid formation, laminin of the laminin
 CC persistence in Alzheimer's disease
 CC laminin products (see W50895) repeats
 CC A or A1 chain, laminin B1 at
 CC laminin G1 chain, the gl

CC the beta-amyloid binding domain of the laminin A chain. A claimed
 CC method for treating an amyloid disease comprises administering a
 CC polypeptide having a conformational similarity to a fragment of a
 CC laminin protein. A method for diagnosing an amyloid disease
 CC involves determining levels of laminin in a sample. Production
 CC for in vivo inhibition of beta-amyloid repeat in vivo provides a method
 CC and methods can be used for the diagnosis, prognosis, monitoring
 CC and treatment of amyloidosis such as Alzheimer's disease, Down's
 CC syndrome and hereditary cerebral haemorrhage with amyloidosis of
 CC the Dutch type (where the specific amyloid is the beta-amyloid
 CC protein), the amyloidosis associated with chronic inflammation,
 CC various forms of malignancy and Familial Mediterranean Fever (AF
 CC amyloid or inflammation-association amyloidosis), the amyloidosis
 CC associated with multiple myeloma and other B-cell abnormalities
 CC (AL amyloid), the amyloidosis associated with type II diabetes
 CC (amylin or islet amyloid), the amyloidosis associated with prion
 CC diseases including Creutzfeldt-Jacob disease, Gerstmann-Strausler
 CC syndrome, kuru and animal scrapie (PrP amyloid), the amyloidosis
 CC associated with long-term haemodialysis and carpal tunnel syndrome
 CC (beta 2-microglobulin amyloid), the amyloidosis associated with
 CC senile cardiac amyloid and familial Amyloidotic Polyneuropathy
 CC (prealbumin or transthyretin amyloid), and the amyloidosis
 CC associated with endocrine tumours such as medullary carcinoma of
 CC the thyroid (variant of procalcitonin).
 SQ Sequence 1801 AA;

Query Match 72.98; Score 43; DB 1; Length 1801;

Best Local Similarity 50.0%; Pred. No. 1.67e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 890 DAHTGACL 897

QY 78 EGTSGCL 85

RESULT 6

ID R05285 standard; protein; 196 AA.

AC R05285;

DT 19-AUG-1990 (first entry)

DE Amino acid sequence of human alpha foetoprotein (hAFP) domain 1

OS Human alpha foetoprotein (hAFP) domain 1; immunosuppressant.

PN J02005856-A.

PD 10-JAN-1990.

PF 27-JUN-1988; 158596.

PR 27-JUN-1988; JP-138596.

PA (KWAN-) Kwansai Shingijutsu.

DR WPI; 90-053912/08.

DR N-PSDB; Q01869.

PT Human alpha foetoprotein domain 1 gene -

including corresp. plasmid recombinant

PS Claim 1; Page 491; 16pp; Japanese.

CC Also claimed in the patent are a recombinant plasmid

CC contg. its DNA, a host cell transformed by this plasmid, and prodn. of it

CC by culturing the transformed cell. A large amt. of it can be prepd.,

CC which is very pure and can be used to assay for development of

CC immunosuppressant. Also specific antibodies against it can be prepd.

SQ Sequence 196 AA;

Query Match 71.2%; Score 42; DB 1; Length 196;

Best Local Similarity 50.0%; Pred. No. 2.18e+02;

Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 61 DEQSSGCL 68

QY 78 EGTSGCL 85

RESULT 7

ID R99228 standard; Protein; 197 AA.

AC R99228;

DT 27-NOV-1996 (first entry)

DE Recombinant human alpha-foetoprotein (aal-197).
 KW Alpha-foetoprotein; AFP; cell proliferation; bone marrow;
 KW autoimmune disease; breast cancer; prostate cancer; neoplasm;
 KW tumour; myelotoxicity; therapy; cell culture medium.
 OS Homo sapiens.

PN WO9622787-A1.

PD 01-AUG-1996.

PF 24-JAN-1996; U00996.

PR 24-JAN-1995; US-377309.

PR 24-JAN-1995; US-377316.

PR 24-JAN-1995; US-377311.

PR 24-JAN-1995; US-377317.

PR 21-JUL-1995; US-505012.

PA (MURG/) MURGITA R A.

PI Murgita RA;

DR WPI; 96-362459/36.

PT New isolated recombinant human alpha-fetoprotein - used for treating
 PT autoimmune diseases or neoplasms, for inhibiting myelotoxicity or
 PT promoting bone marrow cell proliferation.
 PS Example; Page 92; 133pp; English.

CC A recombinant human alpha-foetoprotein (hAFP) fragment (R99228)

CC comprises domain I (amino acids 1-197) of the mature hAFP

CC (see also R99227). It can be obtd. by PCR amplification of hAFP

CC cDNA cloned in p118 using primers Domi25 (T35179) and Domi3

CC (T35180), and expression of the amplified fragment in a prokaryotic

CC (e.g. Escherichia coli) or eukaryotic (e.g. Spodoptera frugiperda)

CC host. Recombinant hAFP fragments (see also R99222-26) can be

CC used e.g. to inhibit autoreactive immune cell proliferation, treat

CC a neoplasm (esp. breast or prostate tumour) and as a cell culture

CC medium component.

SQ Sequence 197 AA;

Query Match 71.2%; Score 42; DB 1; Length 197;

Best Local Similarity 50.0%; Pred. No. 2.18e+02;

Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 61 DEQSSGCL 68

QY 78 EGTSGCL 85

RESULT 8

ID W55731 standard; Protein; 340 AA.

AC W55731;

DT 13-JUL-1998 (first entry)

DE H. pylori ORF 14cpl1908.24218954.cl_68 cell envelope OMP.

KW Cytoplasmic; vaccine; prevention; infection; treatment; envelope;

KW identification; binding compound; bacteria; life cycle; activator;

KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;

KW cell envelope outer membrane protein; OMP; tyrosine cluster motif.

OS Helicobacter pylori.

PN WO9737044-A1.

PD 09-OCT-1997.

PF 27-MAR-1997; U05223;

PR 06-DEC-1996; US-761318.

PR 29-MAR-1996; US-625811.

PR 02-APR-1996; US-758731.

PR 25-OCT-1996; US-736905.

PR 28-OCT-1996; US-738859.

PA (ASTR) ASTRA AB.

PI Alm RA, Smith D;

DR WPI; 97-503122/46.

DR N-PSDB; V25140.

PT Helicobacter pylori nucleic acid sequences and encoded

PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori

PT infection and for diagnosis of H. pylori infection

PS Claims 14,80; Page 1002,1003; 1149pp; English.

CC This sequence is a H. pylori cell envelope outer membrane protein

CC having a terminal Phe residue and a C-terminal tyrosine cluster motif.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds, The

CC useful as potential H. pylori life cycle activators or inhibitors. The

CC DNA and probes derived from it may be used for the identification of

Abstract

33

yke.

51

CC (W12716-23) are involved in the biosynthesis, and regulation of
 CC biosynthesis, of Phl. Bacteria expressing these proteins can be
 CC used in the biological control of fungal pathogens.
 SQ Sequence 389 AA;

Query Match 71.2%; Score 42; DB 1; Length 389;
 Best Local Similarity 83.3%; Pred. No. 2.18e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 346 GHASGC 351
 QY 79 GHTSGC 84

RESULT 12
 ID W01023 standard; Protein; 590 AA.

AC W01023;
 DT 31-MAY-1997 (first entry)
 DE Alpha-foetoprotein.
 KW AFP; alpha-foetoprotein; recombinant production; primer; specific;
 KW contamination risk; polymerase chain reaction.
 OS Homo sapiens.
 PN DE1909169-A1.
 PD 19-SEP-1996.
 PF 14-MAR-1995; DE-009169.
 PR 14-MAR-1995; DE-009169.
 PA (AIRC-) AIR CLIN SERVICE MEDIZIN-GERAETE-VERTRIE.
 PA (ACSA-) ACS AIR CLIN SERVICE MEDIXINE GERAETE VER.
 PI Hoeft H, Kotschneva G, Serpinski O, Sivolobova G;
 PI Tatkov S, Tschishikow V, Uрманov I;
 DR WPI; 96-425952/43.
 DR N-PSDB; T38240, T65016.
 PT Oligo-nucleotide primers for cloning DNA - esp. that encoding
 PT alpha-foetoprotein for expression in mammalian cells
 PS Claim 6; Fig 3; 7pp; German.
 CC The present sequence is that of human alpha-foetoprotein (AFP). It is
 CC encoded by T38240 or T65016. The DNA sequences are useful for the
 CC subsequent expression of AFP in mammalian cells. Recombinant AFP can
 CC replace the natural protein, eliminating the risk of contamination by
 CC pathogens, e.g. for immunodiagnosis. Specific primers (T38236-39) for
 CC AFP DNA are claimed and provide specific amplification of AFP cDNA
 CC (contrast use of non-specific polyt primers) so improve both cDNA
 CC synthesis and PCR.
 SQ Sequence 590 AA;

Query Match 71.2%; Score 42; DB 1; Length 590;
 Best Local Similarity 50.0%; Pred. No. 2.18e+02;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 61 DEOSSGCL 68
 QY 78 EGHSGCL 85

RESULT 13
 ID R99227 standard; Protein; 609 AA.

AC R99227;
 DT 27-NOV-1996 (first entry)
 DE Human alpha-foetoprotein.
 KW Alpha-foetoprotein; AFP; cell proliferation; bone marrow;
 KW autoimmune disease; breast cancer; prostate cancer; neoplasm;
 KW tumour; myelotoxicity; therapy; cell culture medium.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT peptide 1..19
 FT protein /label= Sig_peptide
 FT protein 20..609
 FT protein /label= Mat_protein
 FT domain 20..216
 FT domain /label= Domain-I
 FT domain 217..408
 FT domain /label= Domain-II
 FT domain 409..609

FT W09622787-A1.
 PN 01-AUG-1996.
 PF 24-JAN-1995; U00996.
 PR 24-JAN-1995; US-377309.
 PR 24-JAN-1995; US-377316.
 PR 24-JAN-1995; US-377311.
 PR 24-JAN-1995; US-377317.
 PR 21-JUL-1995; US-505012.
 PA (MURG/) MURGITA R A.
 PI Murgita RA;
 DR WPI; 96-362459/36.
 DR N-PSDB; T35173.

PT New isolated recombinant human alpha-fetoprotein - used for treating
 PT autoimmune diseases or neoplasms, for inhibiting myelotoxicity or
 PT promoting bone marrow cell proliferation.
 PS Example; Fig 1; 133pp; English.
 CC The amino acid sequence (R99227) of full-length human alpha-
 CC foetoprotein (hAFP) was deduced from a cDNA clone (T35173) obtd.
 CC from a human foetal liver cDNA library. Recombinant hAFP can be
 CC produced using standard DNA technology and expressed in either
 CC prokaryotic (esp. Escherichia coli) or eukaryotic (esp. Spodoptera
 CC frugiperda) host cells. Recombinant mature hAFP or its fragments
 CC (see also R99222-28) can be used e.g. to inhibit autoreactive
 CC immune cell proliferation, treat a neoplasm (esp. breast or prostate
 CC tumour) and as a cell culture medium component.
 SQ Sequence 609 AA;

Query Match 71.2%; Score 42; DB 1; Length 609;
 Best Local Similarity 50.0%; Pred. No. 2.18e+02;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 80 DEOSSGCL 87
 QY 78 EGHSGCL 85

RESULT 14

ID R91385 standard; Peptide; 28 AA.
 AC R91385;
 DT 10-OCT-1996 (first entry)
 DE MEH1 library derived peptide, 950406-1.
 DE Cis-element; ligand; highly specific DNA binding domain; HSDB; human;
 KW H2-kappa-B; regulatory domain; murine; homologue; IL-6; interleukin;
 KW HLA class I gene; HLA-B27; HLA-A2; acute phase protein response;
 KW syngene; immune response; IL-8; IL-16; diagnosis; gene therapy;
 KW transcription.
 OS Synthetic.
 PN W09606188-A1.
 PD 29-FEB-1996.
 PF 17-AUG-1995; U10523.
 PR 18-AUG-1994; US-292902.
 PR 15-AUG-1995; US-515190.
 PA (CYTO-) CYTOGEN CORP.
 PI Fowlkes DM, Rodwell JD;
 DR WPI; 96-151391/15.
 PT Synthetic random nucleotide sequences encoding ligand binding
 PT domains - identified by screening library of vectors or peptide(s)
 PT and useful for gene therapy and diagnosis
 PS Claim 70; Page 196; 224pp; English.
 CC The sequences given in R91378-414 represent peptides that bind to
 CC highly specific DNA binding domains (HSDB)'s. These sequences were
 CC tested for binding to the H2-kappa-B oligonucleotide which contains
 CC the NF-kappa-B binding site, and comprises the sequence given in T13579.
 CC These peptides may be used in a composition for diagnosis and gene
 CC therapy and for modifying the transcription or activity of a gene.
 SQ Sequence 28 AA;

Query Match 69.5%; Score 41; DB 1; Length 28;
 Best Local Similarity 57.1%; Pred. No. 2.84e+02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 12 GHATGCF 18

QY 79 GHTSGCL 85

RESULT 15

ID W19761 standard; Protein; 328 AA.
 AC W19761;
 DT 17-SEP-1997 (first entry)
 DE Pyruvate:ferredoxin oxidoreductase subunit (ORF4 product).
 KW Pyruvate:ferredoxin oxidoreductase; ulcer; vaccine; therapy.
 OS Helicobacter pylori.
 PN W09723626-A1.
 PD 03-JUL-1997.
 PF 17-DEC-1996; G03119.
 PR 22-DEC-1995; GB-026407.
 PA (GLAXO) GLAXO GROUP LTD.
 PI Chalk PA, Clayton CL, Hughes NJ, Kelly DJ;
 DR WPI; 97-351063/32.
 DT N-PSDB; T72719.
 PT New isolated Helicobacter pylori oxidoreductase enzymes - used to
 develop products for the diagnosis, treatment and prevention of H.
 PT Pylori mediated diseases or disorders
 PS Example B; Fig 2; 36pp; English.
 CC 4 Polypeptides (W19758-61) comprise the subunits of a newly
 identified tetrameric pyruvate:ferredoxin oxidoreductase (PFOR)
 of Helicobacter pylori. Their amino acid sequences were deduced
 from the open reading frames of a single operon (172719). The
 CC ORF4 product has a predicted mol.wt. of 36.5 kDa and shows sequence
 similarity to Anabaena and Klebsiella pyruvate oxidoreductase. The
 CC PFOR subunit polypeptides can be expressed in transformed host
 CC cells. PFOR and newly identified H. pylori 2-oxoglutarate:acceptor
 CC oxidoreductase (see also W19754-57) are important or essential to H.
 CC pylori but are not expressed by humans and can therefore be used in
 CC vaccines or to screen for agents that can be used to treat H. pylori
 CC diseases or disorders.
 SQ Sequence 328 AA;

Query Match 69.5%; Score 41; DB 1; Length 328;
 Best Local Similarity 57.1%; Pred.No. 2.84e+02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 48 GNTSGCL 54
 QY 79 GHTSGCL 85

Search completed: Wed May 10 12:42:36 2000
 Job time : 7 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```
>US-09-376-430-2
Title:
Description: (78-85) from US09376430A, pep (12 of 25)
Perfect Score: 59
Sequence: 1 EGHTSGCL 8
```

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

```
Database:
a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1
```

Statistics: Mean 13.954; Variance 40.531; scale 0.344

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description	Pred. No.
		Match						
1	45	76.3	1940	2	US-08-644-	Sequence 30,	Applicati	5.93e+01
2	44	74.6	1239	3	US-08-937-	Sequence 2,	Applicatio	7.73e-01
3	42	71.2	197	3	PCT-US96-0	Sequence 6,	Applicatio	1.31e+02
4	42	71.2	197	2	US-08-377-	Sequence 3,	Applicatio	1.31e+02
5	42	71.2	389	3	PCT-US96-1	Sequence 10,	Applicatio	1.31e+02
6	42	71.2	389	3	PCT-US96-0	Sequence 9,	Applicatio	1.31e+02
7	42	71.2	389	2	US-08-377-	Sequence 6,	Applicatio	1.31e+02
8	42	71.2	389	2	US-08-494-	Sequence 10,	Applicatio	1.31e+02
9	42	71.2	590	3	PCT-US96-0	Sequence 5,	Applicatio	1.31e+02
10	42	71.2	590	2	US-08-377-	Sequence 2,	Applicatio	1.31e+02
11	42	71.2	609	3	PCT-US95-0	Sequence 4,	Applicatio	1.31e+02
12	42	71.2	609	1	US-08-222-	Sequence 4,	Applicatio	1.31e+02
13	40	67.8	194	2	US-08-739-	Sequence 8,	Applicatio	2.19e+02
14	40	67.8	194	1	US-08-616-	Sequence 11,	Applicatio	2.19e+02
15	40	67.8	273	1	US-08-152-	Sequence 30,	Applicati	2.19e+02
16	40	67.8	436	2	US-08-846-	Sequence 72,	Applicati	2.19e+02
17	40	67.8	436	2	US-08-846-	Sequence 3,	Applicatio	2.19e+02
18	40	67.8	459	1	US-08-403-	Sequence 5,	Applicatio	2.19e+02
19	39	66.1	192	3	PCT-US95-1	Sequence 66,	Applicati	2.83e+02
20	39	66.1	192	1	US-08-086-	Sequence 66,	Applicati	2.83e+02
21	39	66.1	192	2	US-08-488-	Sequence 66,	Applicati	2.83e+02
22	39	66.1	192	2	US-08-290-	Sequence 66,	Applicati	2.83e+02
23	39	66.1	219	2	US-08-460-	Sequence 18,	Applicati	2.83e+02

CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC FEATURE:
CC NAME/KEY: Rat Agrin
CC LOCATION: 1...1940
CC OTHER INFORMATION:
SQ SEQUENCE 1940 AA; 206559 MW; 18912626 CN;

Query Match 76.3%; Score 45; DB 2; Length 1940;
Best Local Similarity 50.0%; Pred. No. 5.93e+01;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 159 DQOTASCL 166
QY 78 EIGHTSGCL 85

RESULT 2
ID US-08-937-931-2 STANDARD; PRT; 1239 AA.
XX
AC xxxxxx

Sequence 2, Application US/08937931

Sequence 2, Application US/08937931

Patent No. 5935792

GENERAL INFORMATION:

APPLICANT: Rubin, Gerald M.

APPLICANT: Pan, Duojia

APPLICANT: Rooke, Jenny

APPLICANT: Yavari, Reza

APPLICANT: Xu, Tian

TITLE OF INVENTION: KUZ: A NO. 5935792el Family of Metalloproteases

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

STREET: 268 BUSH STREET, SUITE 3200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/937,931

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: OSMAN, RICHARD A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: B97-081

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 343-4341

TELEFAX: (415) 343-4342

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1239 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE 1239 AA; 136522 MW; 7567283 CN;

Query Match 74.6%; Score 44; DB 2; Length 1239;
Best Local Similarity 71.4%; Pred. No. 7.73e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 273 EGHVAGC 279
QY 78 EGHISGC 84

RESULT 3
ID PCT-US96-00996-6 STANDARD; PRT; 197 AA.
XX
AC xxxxxx

Sequence 6, Application PC/TUS9600996

Sequence 6, Application PC/TUS9600996

GENERAL INFORMATION:

APPLICANT: Murgita, Robert A.

TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF CLONED

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street, Suite 3100

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/00996

FILING DATE: 24-JAN-1996

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/377,317

FILING DATE: 24-JAN-1995

CLASSIFICATION:

FILING DATE: 24-JAN-1995

APPLICATION NUMBER: 08/377,311

CLASSIFICATION:

FILING DATE: 24-JAN-1995

APPLICATION NUMBER: 08/377,316

CLASSIFICATION:

FILING DATE: 24-JAN-1995

APPLICATION NUMBER: 08/505,012

CLASSIFICATION:

FILING DATE: 21-JULY-1995

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.

REGISTRATION NUMBER: 30,162

REFERENCE/DOCKET NUMBER: 06727/003001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 197 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 197 AA; 22233 MW; 192091 CN;

Query Match 71.2%; Score 42; DB 3; Length 197;
Best Local Similarity 50.0%; Pred. No. 1.31e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 61 DEOSSGCL 68
QY 78 EGTSGCL 85

RESULT 4
ID US-08-377-309-3 STANDARD; PRT; 197 AA.
XX XXXXX

AC
XX
DT

Sequence 3, Application US/08377309A

Sequence 3, Application US/08377309A

Patent No. 5965528

GENERAL INFORMATION:

APPLICANT: Murgita, Robert A.

TITLE OF INVENTION: RECOMBINANT HUMAN ALPHA-FETOPROTEIN AS

TITLE OF INVENTION: AN IMMUNOSUPPRESSIVE AGENT

FILE REFERENCE: 06727/005001

CURRENT APPLICATION NUMBER: US/08/377,309A

CURRENT FILING DATE: 1995-01-24

NUMBER OF SEQ ID NOS: 16

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 3

LENGTH: 197

TYPE: PRT

ORGANISM: Homo sapiens

SEQUENCE 197 AA; 22233 MW; 192091 CN;

Query Match 71.2%; Score 42; DB 2; Length 197;

Best Local Similarity 50.0%; Pred. No. 1.31e+02;

Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 61 DEOSSGCL 68
QY 78 EGTSGCL 85

RESULT 5
ID PCT-US96-10986-10 STANDARD; PRT; 389 AA.
XX XXXXX

AC
XX
DT

Sequence 10, Application PC/TUS9610986

Sequence 10, Application PC/TUS9610986

GENERAL INFORMATION:

TITLE OF INVENTION: Sequences for Production of

TITLE OF INVENTION: 2,4-Diacetylphloroglucinol and Methods

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: Stephan A. Pendorf, DOMINIK & STEIN

STREET: 600 N. West Shore Boulevard, Suite 1000

CITY: Tampa

STATE: FL

COUNTRY: USA

ZIP: 33609

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/10986

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Pendorf, Stephan A.

REGISTRATION NUMBER: 32665

CC REFERENCE/DOCKET NUMBER: A700.320

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (813) 289-2966

CC TELEFAX: (813) 289-2967

CC INFORMATION FOR SEQ ID NO: 10:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 389 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 389 AA; 41721 MW; 724304 CN;

Query Match 71.2%; Score 42; DB 3; Length 389;

Best Local Similarity 83.3%; Pred. No. 1.31e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 346 GHASGC 351

QY 79 GHTSGC 84

|||||

RESULT 6

ID PCT-US96-00996-9 STANDARD; PRT; 389 AA.

XX XXXXX

AC

XX

DT

XX

DE

XX

Sequence 9, Application PC/TUS9600996

Sequence 9, Application PC/TUS9600996

GENERAL INFORMATION:

APPLICANT: Murgita, Robert A.

TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF CLONED

TITLE OF INVENTION: HUMAN ALPHA-FETOPROTEIN

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street, Suite 3100

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/00996

FILING DATE: 24-JAN-1996

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/377,317

FILING DATE: 24-JAN-1995

CLASSIFICATION:

APPLICATION NUMBER: 08/377,311

FILING DATE: 24-JAN-1995

CLASSIFICATION:

APPLICATION NUMBER: 08/377,309

FILING DATE: 24-JAN-1995

CLASSIFICATION:

APPLICATION NUMBER: 08/377,316

FILING DATE: 24-JAN-1995

CLASSIFICATION:

APPLICATION NUMBER: 08/505,012

FILING DATE: 21-JULY-1995

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.

REGISTRATION NUMBER: 30,162

REFERENCE/DOCKET NUMBER: 06727/003001

TELECOMMUNICATION INFORMATION:

CC	TELEPHONE: (617) 542-5070	
CC	TELEFAX: (617) 542-8906	
CC	TELEX: 200154	
CC	INFORMATION FOR SEQ ID NO: 9:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH: 389 amino acids	
CC	TYPE: amino acid	
CC	STRANDEDNESS: not relevant	
CC	TOPOLOGY: linear	
CC	MOLECULE TYPE: protein	
CC	SEQUENCE 389 AA; 44231 MW; 742166 CN;	
Db	61 DEOSSGCL 68	
Qy	78 EGTSGCL 85	
RESULT	7	
ID	US-08-377-309-6	STANDARD; PRT; 389 AA.
XX		
AC	xxxxxx	
XX		
DT		
XX		
XX		
DE	Sequence 6, Application US/08377309A	
XX		
CC	Sequence 6, Application US/08377309A	
CC	Patent No. 5965528	
CC	GENERAL INFORMATION:	
CC	APPLICANT: Murgita, Robert A.	
CC	TITLE OF INVENTION: RECOMBINANT HUMAN ALPHA-FETOPROTEIN AS	
CC	TITLE OF INVENTION: AN IMMUNOSUPPRESSIVE AGENT	
CC	FILE REFERENCE: 06727/005001	
CC	CURRENT APPLICATION NUMBER: US/08/377,309A	
CC	CURRENT FILING DATE: 1995-01-24	
CC	NUMBER OF SEQ ID NOS: 16	
CC	SOFTWARE: FastSeq for Windows Version 3.0	
CC	SEQ ID NO 6	
CC	LENGTH: 389	
CC	TYPE: PRT	
CC	ORGANISM: Homo sapiens	
CC	SEQUENCE 389 AA; 44231 MW; 742166 CN;	
SQ		
Query Match	71.2%; Score 42; DB 2; Length 389;	
Best Local Similarity	50.0%; Pred. No. 1.31e+02;	
Matches	4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;	
Db	61 DEOSSGCL 68	
Qy	78 EGTSGCL 85	
RESULT	8	
ID	US-08-494-907-10	STANDARD; PRT; 389 AA.
XX		
AC	xxxxxx	
XX		
DT		
XX		
DE	Sequence 10, Application US/08494907	
XX		
CC	Sequence 10, Application US/08494907	
CC	Patent No. 5955298	
CC	GENERAL INFORMATION:	
CC	APPLICANT: Thomashow, Linda S	
CC	APPLICANT: Bangera, Mahalaxmi	
CC	APPLICANT: Weller, David M	
CC	APPLICANT: Cook, R. James	
CC	TITLE OF INVENTION: Sequences for Production of	

CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: 08/377,317
 CC FILING DATE: 24-JAN-1995
 CC CLASSIFICATION:
 CC APPLICATION NUMBER: 08/377,311
 CC FILING DATE: 24-JAN-1995
 CC CLASSIFICATION:
 CC APPLICATION NUMBER: 08/377,309
 CC FILING DATE: 24-JAN-1995
 CC CLASSIFICATION:
 CC APPLICATION NUMBER: 08/377,316
 CC FILING DATE: 24-JAN-1995
 CC CLASSIFICATION:
 CC APPLICATION NUMBER: 08/505,012
 CC FILING DATE: 21-JULY-1995
 CC CLASSIFICATION:
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Clark, Paul T.
 CC REGISTRATION NUMBER: 30,162
 CC REFERENCE/DOCKET NUMBER: 06727/003001
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (617) 542-5070
 CC TELEFAX: (617) 542-8906
 CC TELEX: 200154
 CC INFORMATION FOR SEQ ID NO: 5:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 590 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: not relevant
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC SEQUENCE 590 AA; 66278 MW; 1678098 CN;

Query Match 71.2%; Score 42; DB 3; Length 590;
 Best Local Similarity 50.0%; Pred. No. 1.31e+02;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 61 DEQSSGCL 68
 : : : : :
 QY 78 EGTSGCL 85

RESULT 10
 ID US-08-377-309-2 STANDARD; PRT; 590 AA.
 XX
 AC xxxxxx
 XX
 DT
 DE
 XX

Sequence 2, Application US/08377309A
 XX
 XX Sequence 2, Application US/08377309A
 CC Patent No. 5965528
 CC GENERAL INFORMATION:

CC APPLICANT: Murgita, Robert A.
 CC TITLE OF INVENTION: RECOMBINANT HUMAN ALPHA-FETOPROTEIN AS
 CC FILE REFERENCE: 06727/005001
 CC CURRENT APPLICATION NUMBER: US/08/377,309A
 CC CURRENT FILING DATE: 1995-01-24
 CC NUMBER OF SEQ ID NOS: 16
 CC SOFTWARE: FastSeq for Windows Version 3.0
 CC SEQ ID NO 2
 CC LENGTH: 590
 CC TYPE: PRT
 CC ORGANISM: Homo sapiens
 CC SEQUENCE 590 AA; 66278 MW; 1678098 CN;

Query Match 71.2%; Score 42; DB 2; Length 590;
 Best Local Similarity 50.0%; Pred. No. 1.31e+02;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 61 DEQSSGCL 68

QY 78 EGTSGCL 85
 : : : : :
 RESULT 11

ID PCT-US95-04075-4 STANDARD; PRT; 609 AA.
 XX
 AC xxxxxx
 XX
 DT
 DE

Sequence 4, Application PC/TUS9504075

XX
 CC Sequence 4, Application PC/TUS9504075
 CC GENERAL INFORMATION:
 CC APPLICANT: AMGEN INC.
 CC TITLE OF INVENTION: Afamin: A Human Serum Albumin-Like
 CC TITLE OF INVENTION: Protein
 CC NUMBER OF SEQUENCES: 33
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Amgen Center, Patent Operations/RRC
 CC STREET: 1840 DeHavilland Drive
 CC CITY: Thousand Oaks
 CC STATE: California
 CC COUNTRY: U.S.
 CC ZIP: 91320-1789
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: PCT/US95/04075
 CC FILING DATE:
 CC CLASSIFICATION:
 CC INFORMATION FOR SEQ ID NO: 4:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 609 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: unknown
 CC TOPOLOGY: unknown
 CC MOLECULE TYPE: protein
 CC SEQUENCE 609 AA; 68677 MW; 1795387 CN;

Query Match 71.2%; Score 42; DB 3; Length 609;
 Best Local Similarity 50.0%; Pred. No. 1.31e+02;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 80 DEQSSGCL 87
 : : : : :
 QY 78 EGTSGCL 85

RESULT 12
 ID US-08-222-619-4 STANDARD; PRT; 609 AA.
 XX
 AC xxxxxx
 XX
 DT
 DE

Sequence 4, Application US/08222619

XX
 CC Sequence 4, Application US/08222619
 CC Patent No. 5652352
 CC GENERAL INFORMATION:
 CC APPLICANT: Lichenstein, Henri
 CC APPLICANT: Lyons, David
 CC APPLICANT: Wurfel, Mark
 CC APPLICANT: Wright, Samuel
 CC TITLE OF INVENTION: Afamin: A Human Serum Albumin-Like
 CC TITLE OF INVENTION: Protein
 CC NUMBER OF SEQUENCES: 33
 CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Amgen Center, Patent Operations/RRC
CC STREET: 1840 DeHavilland Drive
CC CITY: Thousand Oaks
CC STATE: California
CC COUNTRY: U.S.
CC ZIP: 91320-1789
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/222,619
CC FILING DATE:
CC CLASSIFICATION: 435
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 609 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC SEQUENCE 609 AA; 68677 MW; 1795387 CN;

Query Match 71.2%; Score 42; DB 1; Length 609;
Best Local Similarity 50.0%; Pred. No. 1.31e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 80 DEOSSGCL 87
QY 78 EGTSGCL 85

RESULT 13
ID US-08-739-485-8 STANDARD; PRT; 194 AA.
XX xxxxxx
Sequence 8, Application US/08739485
Sequence 8, Application US/08739485
Patent No. 5863898
GENERAL INFORMATION:
APPLICANT: Goli, Surya K.
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/739,485
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0142 US

CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-855-0555
CC TELEFAX: 415-845-4166
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 194 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC LIBRARY: GenBank
CC CLONE: 1234841
CC SEQUENCE 194 AA; 20969 MW; 197881 CN;

Query Match 67.8%; Score 40; DB 2; Length 194;
Best Local Similarity 57.1%; Pred. No. 2.19e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 74 GOGAGCL 80
QY 79 GHTSGCL 85

RESULT 14
ID US-08-616-368A-11 STANDARD; PRT; 194 AA.
XX xxxxxx
Sequence 11, Application US/08616368A
Sequence 11, Application US/08616368A
Patent No. 5767262
GENERAL INFORMATION:
APPLICANT: Lee, Mu-En
APPLICANT: Haber, Edgar
APPLICANT: Jain, Mukesh
APPLICANT: Yet, Shaw-Fang
TITLE OF INVENTION: SMOOTH MUSCLE CELL LIM PROTEIN
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/616,368A
FILING DATE: 15-MAR-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/022001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein

W P S R L H
(TK)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 12:42:03 2000; MasPar time 3.91 Seconds
Tabular output not generated. 96.543 Million cell updates/sec

Title: >US-09-376-430-2
Description: (78-85) from US09376430A.pep (12 of 25)
Perfect Score: 59
Sequence: 1 EIGHTSGCL 8

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r2
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 19.788; Variance 23.761; scale 0.833

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	51	86.4	236	B70856	hypothetical protein	2.56e+01
2	47	79.7	204	C69468	conserved hypothetical	2.11e+00
3	45	76.3	1959	AGRT	agrin - rat	5.84e+00
4	45	76.3	1966	T08991	hypothetical protein	5.84e+00
5	44	74.6	205	D69036	conserved hypothetical	9.63e+00
6	44	74.6	1239	T13809	probable disintegrin	9.63e+00
7	44	74.6	1955	AGCH	agrin precursor - chi	9.63e+00
8	43	72.9	902	B40494	androgen receptor - r	1.57e+01
9	43	72.9	1244	S37034	DNA-directed DNA poly	1.57e+01
10	43	72.9	1801	MMRTS	laminin beta-2 chain	1.57e+01
11	42	71.2	156	C65065	prelamin peptidase de	2.56e+01
12	42	71.2	205	G71891	hypothetical protein	2.56e+01
13	42	71.2	205	E64621	conserved hypothetical	2.56e+01
14	42	71.2	329	AJ2302	glutamate--ammonia li	2.56e+01
15	42	71.2	343	AJSM2V	glutamate--ammonia li	2.56e+01
16	42	71.2	388	C71813	probable outer membra	2.56e+01
17	42	71.2	388	E64707	outer membrane protei	2.56e+01
18	42	71.2	609	FPHU	alpha-fetoprotein pre	2.56e+01
19	42	71.2	1309	T00078	probable RNA-directed	2.56e+01
20	41	69.5	125	A70020	hypothetical protein	4.12e+01
21	41	69.5	171	G70636	hypothetical protein	4.12e+01
22	41	69.5	261	A34476	collagen alpha 2(IV)	4.12e+01
23	41	69.5	278	C69113	ribokinase - Methanob	4.12e+01

24 41 69.5 314 1 G64558 probable pyruvate syn 4.12e+01
25 41 69.5 314 2 A71858 pyruvate ferredoxin o 4.12e+01
26 41 69.5 406 2 T06666 probable proteasome r 4.12e+01
27 41 69.5 438 1 XXMSN phosphatidylcholine-- 4.12e+01
28 41 69.5 440 1 XXRTN phosphatidylcholine-- 4.12e+01
29 41 69.5 440 1 XXHUN phosphatidylcholine-- 4.12e+01
30 41 69.5 440 2 JC1502 phosphatidylcholine-- 4.12e+01
31 41 69.5 444 2 T04812 bindin precursor - se 4.12e+01
32 41 69.5 480 2 S14394 hypohetical protein 4.12e+01
33 41 69.5 523 2 T04825 probable methylmalona 4.12e+01
34 41 69.5 532 2 T02721 tyrosine phosphoprote 4.12e+01
35 41 69.5 533 2 B56110 tyrosinethetical prote 4.12e+01
36 41 69.5 906 2 T00039 laminin gamma-1 chain 4.12e+01
37 41 69.5 1639 1 MMFFB2 collagen alpha 2(IV) 4.12e+01
38 41 69.5 1763 2 S16366 genome polyprotein - 4.12e+01
39 41 69.5 2344 1 RRWRH genome polyprotein - 4.12e+01
40 41 69.5 2344 2 S55399 genome polyprotein - 4.12e+01
41 41 69.5 2344 2 S64740 genome polyprotein - 4.12e+01
42 40 67.8 194 2 S57472 murine muscle LIM pro 6.57e+01
43 40 67.8 194 2 A55099 muscle LIM protein - 6.57e+01
44 40 67.8 584 4 VCHUER retrovirus-related en 6.57e+01
45 40 67.8 1387 2 JC5502 G-protein signaling r 6.57e+01

ALIGNMENTS

RESULT 1
ENTRY B70856 #type complete
TITLE hypothetical protein Rv3007c - Mycobacterium tuberculosis (strain H37RV)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
ACCESSION B70856
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
#cross-references MUID:98295987
#accession B70856
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-204 #label COL
#cross-references GB:AU021287; GB:AL123456; NID:g3261508; PID:e1237752; PID:g2791604
#experimental_source strain H37RV
GENETICS
#gene Rv3007C
SUMMARY #length 204 #molecular-weight 22587 #checksum 1968
Query Match 86.4%; Score 51; DB 2; Length 204;
Best Local Similarity 62.5%; Pred. No. 2.56e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 37 .dGHPAGCI; 44
QY 78 EIGHTSGCL 85

RESULT 2
ENTRY C69468 #type complete
TITLE conserved hypothetical protein AF1748 - Archaeoglobus

```

#formal_name Archaeoglobus fulgidus
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
05-Jun-1998
C69468
A69250
#authors
Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
Peterson, J.D.; Richardson, D.L.; Kervilavage, A.R.; Graham,
D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
Peterson, S.; Reich, C.I.; McNeill, L.K.; Badger, J.H.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
C.R.; Venter, J.C.
#journal
Nature (1997) 390:364-370
#title
The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession
C69468
#status
Preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues
1-236 ##label KLE
##cross-references GB:AE000982; GB:AE000782; NID:g2689305; PID:g2648803;
TIGR:AF1748
SUMMARY
#length 236 #molecular-weight 26635 #checksum 2395
Query Match 79.7%; Score 47; DB 2; Length 236;
Best Local Similarity 85.7%; Pred. No. 2,11e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 140 GHTSGCI 146
Qy 79 GHTSGCL 85
RESULT 3
ENTRY #type complete
agrin - rat
#formal_name Rattus norvegicus #common_name Norway rat
31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change
22-Jun-1999
JH0399; A38856
#authors
Rupp, F.; Payan, D.G.; Magill-Solc, C.; Cowan, D.M.;
Schellier, R.H.
#journal
Neuron (1991) 6:811-823
#title
Structure and expression of a rat agrin.
#cross-references MUID:91222570
#accession
JH0399
##molecule_type mRNA
##residues
1-1779;1799-1959 #label RUP
#cross-references GB:M64780; NID:g202798; PIDN:AAA40703.1; PID:g202800
#experimental_source embryonic spinal cord
#note
it is uncertain whether Met-1, Met-18, or Met-24 is the
initiator
#accession
A38856
#authors
Rupp, F.; Oezcelik, T.; Linial, M.; Peterson, K.; Francke,
U.; Schellier, R.
#journal
J. Neurosci. (1992) 12:3535-3544
#title
Structure and chromosomal localization of the mammalian agrin
gene.
#cross-references MUID:92407628
#accession
A38856
##molecule_type mRNA
##residues
1780-1798 #label RU2
#cross-references GB:S44194
COMMENT This protein mediates the motor neuron-induced aggregation of

```

acetylcholine receptors and acetylcholine-esterase on the surface of muscle fibers of the neuromuscular junction.

90% of rat embryonic transcripts encode the variant labeled below as form 3. However, alternative splicing may produce as many as eight different forms of agrin, which differ in their acetylcholine receptor clustering activity.

CLASSIFICATION #superfamily agrin; agrin inhibitor-like repeat homology; EGF homology; Kazal proteinase inhibitor homology; laminin G repeat homology; laminin-type EGF-like homology; alternative splicing; duplication; glycoprotein; neuromuscular junction

KEYWORDS

FEATURE

1-1959 #product agrin, form 1 #status predicted #label AG1\

1-1787,1799-1959 #product agrin, form 4 #status predicted #label AG4\

1-1779,1799-1959 #product agrin, form 3 #status predicted #label AG3\

1-1779,1788-1959 #product agrin, form 5 #status predicted #label AG5\

1-1143,1153-1959 #product agrin, form 2 #status predicted #label AG2\

22-50 #region hydrophobic\

65-137 #domain agrin inhibitor-like repeat homology #label AG11\

88-137 #domain Kazal proteinase inhibitor homology #label KP11\

141-212 #domain agrin inhibitor-like repeat homology #label AG12\

163-212 #domain Kazal proteinase inhibitor homology #label KP12\

213-284 #domain agrin inhibitor-like repeat homology #label AG13\

236-284 #domain Kazal proteinase inhibitor homology #label KP13\

288-356 #domain agrin inhibitor-like repeat homology #label AG14\

307-356 #domain Kazal proteinase inhibitor homology #label KP14\

361-429 #domain agrin inhibitor-like repeat homology #label AG15\

381-429 #domain Kazal proteinase inhibitor homology #label KP15\

430-494 #domain agrin inhibitor-like repeat homology #label AG16\

446-494 #domain Kazal proteinase inhibitor homology #label KP16\

495-559 #domain agrin inhibitor-like repeat homology #label AG17\

511-559 #domain Kazal proteinase inhibitor homology #label KP17\

540-542 #region motor neuron attachment (L-R-E) motif\

575-645 #domain agrin inhibitor-like repeat homology #label AG18\

596-645 #domain Kazal proteinase inhibitor homology #label KP18\

688-739 #domain laminin-type EGF-like homology #label LE1\

742-786 #domain laminin-type EGF-like homology #label LE2\

794-864 #domain agrin inhibitor-like repeat homology #label AG19\

814-864 #domain Kazal proteinase inhibitor homology #label KP19\

869-992 #region serine/threonine-rich\

1084-1086 #region motor neuron attachment (L-R-E) motif\

1147-1215 #region serine/threonine-rich\

1224-1257 #domain EGF homology #label EG1\

1267-1442 #domain laminin G repeat homology #label LG1\

1444-1476 #domain EGF homology #label EG2\

1483-1515 #domain EGF homology #label EG3\

1555-1706 #domain laminin G repeat homology #label LG2\

1713-1747 #domain EGF homology #label EG4\

1807-1959 #domain laminin G repeat homology #label LG3\

97-115,105-137, 171-191,180-212, 244-263,252-284, 316-335,324-356, 389-408,397-429, 454-473,462-494, 518-538,527-559, 604-624,613-645, 823-843,832-864, 1224-1235, 1229-1246, 1248-1257, 1444-1455, 1449-1465, 1467-1476,

```

##status      preliminary; nucleic acid sequence not shown;
              translation not shown
##molecule_type  DNA
##residues      1-205 ##label MTH
##cross-references GB:AE000893; GB:AE000666; NID:g2622375; PID:g2622381
##experimental_source strain Delta H
GENETICS
#gene          MTH1267
CLASSIFICATION #superfamily glyoxalase
SUMMARY        #length 205 #molecular-weight 22044 #checksum 4640
Query Match    74.6%; Score 44; DB 2; Length 205;
Best Local Similarity 71.4%; Pred. No. 9.63e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 128 GHTPGCI 134
   |||:||:
Qy 79 GHTSGCL 85

RESULT 6
ENTRY  T13809 #type complete
TITLE  probable disintegrin (EC 3.4.24.-) - fruit fly (Drosophila
        melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change
      T13809
REFERENCE  T1772
          Rooke, J.; Pan, D.; Xu, T.; Rubin, G.M.
          Science (1996) 273:1227-1231
          Kuz, a conserved metalloprotease-disintegrin with two roles
          during Drosophila neurogenesis.
          T13809
#accession
##status      preliminary; translated from GB/EMBL/DBD
##molecule_type mRNA
##residues    1-1239 ##label ROO
##cross-references EMBL:U60591; NID:g1531632; PID:g1531633;
              PIDN:AAC47275.1
GENETICS
#gene          kuz
#map_position  2
KEYWORDS        hydrolase; metalloproteinase
SUMMARY        #length 1239 #molecular-weight 136522 #checksum 5010
Query Match    74.6%; Score 44; DB 2; Length 1239;
Best Local Similarity 71.4%; Pred. No. 9.63e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 273 EGHVAGC 279
   |||:||:
Qy 78 EGHTSGC 84

RESULT 7
ENTRY  AGCH #type complete
TITLE  agrin precursor - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change
      22-Jun-1999
ACCESSIONS    JH0591; A38857; B38857; I50692
REFERENCE      JH0591
#authors      Tsim, K.W.K.; Ruegg, M.A.; Escher, G.; Kroeger, S.; McMahan,
              U.J.
#journal      Neuron (1992) 8:677-689
#title        cDNA that encodes active agrin.
#cross-references MUID:92232297
#accession    JH0591
#molecule_type mRNA
#residues     1-1955 ##label TSI
#cross-references GB:M94271; NID:g211120; PIDN:AAA48585.1; PID:g211121
#experimental_source brain
REFERENCE      A38857

```

```

#authors      Ruegg, M.A.; Tsim, K.W.K.; Horton, S.E.; Kroeger, S.; Escher,
               G.; Gensch, E.M.; McMahon, U.J.
#journal      Neuron (1992) 8:691-699
#title        The agrin gene codes for a family of basal lamina proteins
               that differ in function and distribution.
#cross-references MUID:92232298
#contents     alternative splicing
#accession    A38857
               ##molecule_type mRNA
               ##residues      1132-1783;1795-1955 ##label RU2
               ##cross-references GB:M97371
               ##molecule_type mRNA
               ##residues      1221-1647;1652-1783;1794-1955 ##label RU3
               ##cross-references GB:M97372
               ##note          translation of the nucleotide sequence is not complete
REFERENCE     I50692
#authors      Thomas, W.S.; O'Dowd, D.K.; Smith, M.A.
#journal      Dev. Biol. (1993) 158:523-535
#title        Developmental expression and alternative splicing of chick
               agrin RNA.
#cross-references MUID:93345745
#accession    I50692
               ##status      preliminary; translated from GB/EMBL/DBJ
               ##molecule_type mRNA
               ##residues      'SHUSNEIPA',1784-1795 ##label THO
               ##cross-references EMBL:U07271; NID:9459665; PIDN:AAA16788.1;
               PID:9459666
COMMENT       This protein mediates the motor neuron-induced aggregation of
               acetylcholine receptors and acetylcholine-esterase on the surface
               of muscle fibers of the neuromuscular junction.
COMMENT       Alternative splicing produces two inactive proteins: agrin-related
               protein 1 and agrin-related protein 2.
CLASSIFICATION superfamily agrin; agrin inhibitor-like repeat homology; EGF
               homology; kazal proteinase inhibitor homology; laminin G
               repeat homology; laminin-type EGF-like homology
KEYWORDS      alternative splicing; duplication; glycoprotein;
               neuromuscular junction
FEATURE       1-38      #domain signal sequence #status predicted #label SIG\
39-1955      #product agrin #status predicted #label MAT\
39-1783,1795-1955  #product agrin-related protein 1 #status predicted
               #label AG1\
39-1647,1652-1783,  #product agrin-related protein 2 #status predicted
1794-1955      #label AG2\
54-126      #domain agrin inhibitor-like repeat homology #label
               AG11\
77-126      #domain kazal proteinase inhibitor homology #label KP11\
130-201      #domain agrin inhibitor-like repeat homology #label
               AG12\
152-201      #domain kazal proteinase inhibitor homology #label KP12\
202-273      #domain agrin inhibitor-like repeat homology #label
               AG13\
225-273      #domain kazal proteinase inhibitor homology #label KP13\
277-344      #domain agrin inhibitor-like repeat homology #label
               AG14\
295-344      #domain kazal proteinase inhibitor homology #label KP14\
350-418      #domain agrin inhibitor-like repeat homology #label
               AG15\
370-418      #domain kazal proteinase inhibitor homology #label KP15\
419-483      #domain agrin inhibitor-like repeat homology #label
               AG16\
435-483      #domain kazal proteinase inhibitor homology #label KP16\
484-548      #domain agrin inhibitor-like repeat homology #label
               AG17\
500-548      #domain kazal proteinase inhibitor homology #label KP17\
563-633      #domain agrin inhibitor-like repeat homology #label
               AG18\
584-633      #domain kazal proteinase inhibitor homology #label KP18\
675-726      #domain laminin-type EGF-like homology #label LE1\
729-773      #domain laminin-type EGF-like homology #label LE2\
781-851      #domain agrin inhibitor-like repeat homology #label
AG19\
#domain kazal proteinase inhibitor homology #label KP19\
#region serine/threonine-rich\
#region serine/threonine-rich\
#domain EGF homology #label EG1\
#domain laminin G repeat homology #label LG1\
#region motor neuron attachment (L-R-E) motif\
#domain EGF homology #label EG2\
#domain EGF homology #label EG3\
#domain laminin G repeat homology #label LG2\
#domain EGF homology #label EG4\
#domain laminin G repeat homology #label LG3\
801-851
856-995
1150-1219
1233-1284
1294-1448
1429-1431
1450-1482
1489-1521
1560-1711
1718-1751
1803-1955
86-105,94-126,
160-180,169-201,
233-252,241-273,
304-323,312-344,
378-397,386-418,
443-462,451-483,
507-527,516-548,
592-612,601-633,
810-830,819-851,
1233-1244,
1238-1253,
1255-1264,
1450-1461,
1455-1471,
1473-1482,
1489-1500,
1494-1510,
1512-1521
390,659,764,814
SUMMARY      #length 1955 #molecular-weight 211411 #checksum 3924
               #disulfide_bonds #status predicted\
               #binding_site carbohydrate (asn) (covalent) #status
               predicted
Query Match 74.6%; Score 44; DB 1; Length 1955;
Best Local Similarity 50.0%; Pred. No. 9.63e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 148 DQTAGCV 155
               :|:|:|:|
QY 78 EGHISGCL 85
RESULT      8
ENTRY       B40494 #type complete
TITLE       androgen receptor - rat
ORGANISM    Rattus norvegicus #common_name Norway rat
DATE        14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change
               20-Sep-1999
ACCESSIONS  B40494; A34943; A36283; B40108
REFERENCE    Chang, C.; Kokontis, J.; Liao, S.
               Proc. Natl. Acad. Sci. U.S.A. (1988) 85:7211-7215
               Structural analysis of complementary DNA and amino acid
               sequences of human and rat androgen receptors.
               #cross-references MUID:89017168
               #accession    B40494
               ##status      preliminary
               ##molecule_type mRNA
               ##residues    1-902 #label CHA
               ##cross-references GB:M33264; NID:g202967; PIDN:AAA40759.1; PID:g202968
               A34943
REFERENCE    Tan, J.; Joseph, D.R.; Quaraby, V.E.; Lubahn, D.B.; Sar, M.;
               French, F.S.; Wilson, E.M.
               Mol. Endocrinol. (1988) 2:1276-1285
               The rat androgen receptor: primary structure, autoregulation
               of its messenger ribonucleic acid, and immunocytochemical
               localization of the receptor protein.
               #cross-references MUID:89112209
               #accession    A34943
               ##molecule_type mRNA
               ##residues    1-388,'S',390-902 ##label TAN
               ##cross-references GB:M20133; NID:g202895; PIDN:AAA40733.1; PID:g202896

```

```

REFERENCE      A36283
#authors      Yarbrough, W.G.; Quarby, V.E.; Simental, J.A.; Joseph, D.R.;
              Sar, M.; Lubahn, D.B.; Olsen, K.L.; French, F.S.; Wilson,
              E.M.
#journal      J. Biol. Chem. (1990) 265:8893-8900
#title       A single base mutation in the androgen receptor gene causes
              androgen insensitivity in the testicular feminized rat.
#cross-references MUID:90256822
#accession   A36283
##molecule_type mRNA
##residues   1-194,196-902 ##label YAR
##cross-references GB:J05454
REFERENCE      A40108
#authors      Chang, C.; Kokontis, J.; Liao, S.
#journal      Science (1988) 240:324-326
#title       Molecular cloning of human and rat complementary DNA encoding
              androgen receptors.
#cross-references MUID:88178111
#accession   B40108
##status     preliminary
##molecule_type mRNA
##residues   540-611 ##label CH2
CLASSIFICATION #superfamily unassigned erba-related proteins; erba
              transforming protein homology
              DNA binding; zinc finger
KEYWORDS       #domain erba transforming protein homology #label ERBA\
540-798        #region zinc finger\
542-562        #region zinc finger
578-602
SUMMARY        length 902 #molecular-weight 98243 #checksum 3069
Query Match   72.9%; Score 43; DB 2; Length 902;
Best Local Similarity 77.8%; Pred. No. 1.57e+01;
Matches 7; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 103 EGPESGCL 111
|||||
QY 78 EGT-SGCL 85

RESULT 9
ENTRY   S37034 #type complete
TITLE   DNA-directed DNA polymerase (EC 2.7.7.7) - African swine
        fever virus
ORGANISM #formal_name African swine fever virus, ASFV
DATE      09-Dec-1993 #sequence_revision 13-Mar-1997 #text_change
18-Jun-1999
ACCESSIONS S37034
REFERENCE   #authors Martins, A.; Costa, J.V.; Ribeiro, G.
#submission submitted to the EMBL Data Library, June 1993
#description Nucleotide sequence of the DNA polymerase gene of African
              swine fever virus.
#accession   S37034
##status     preliminary
##molecule_type DNA
##residues   1-1244 ##label MAR
##cross-references EMBL:X73330; NID:9397585; PIDN:CAA51757.1;
              PID:9397586
CLASSIFICATION #superfamily African swine fever virus DNA-directed DNA
              polymerase
KEYWORDS       nucleotidyltransferase
              #length 1244 #molecular-weight 142533 #checksum 4469
SUMMARY
Query Match   72.9%; Score 43; DB 2; Length 1244;
Best Local Similarity 83.3%; Pred. No. 1.57e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 353 HTPGCL 358
|||||
QY 80 HTSGCL 85

```

```

RESULT 10
ENTRY   MMRTS #type complete
TITLE   laminin beta-2 chain precursor - rat
ALTERNATE_NAMES laminin chain B3; S-laminin
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE      30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change
16-Jul-1999
ACCESSIONS S03539
REFERENCE   #authors Hunter, D.D.; Shah, V.; Merlie, J.P.; Sanes, J.R.
#journal    Nature (1989) 338:229-234
#title     A laminin-like adhesive protein concentrated in the synaptic
              cleft of the neuromuscular junction.
#cross-references MUID:89159410
#accession   S03539
##molecule_type mRNA
##residues   1-1801 ##label HUN
##cross-references EMBL:X16563; NID:957250; PIDN:CAA34561.1; PID:957251
              Laminins are trimers of an alpha-type, a beta-type, and a
              gamma-type laminin chain.
FUNCTION      interact with cells and with other basement membrane proteins
              to promote differentiation, development, and cell migration
              #superfamily laminin beta-1 chain; laminin-type EGF-like
              homology
CLASSIFICATION #domain laminin-type EGF-like homology #label LE01\
              #domain V #label DOM5\
              #domain laminin-type EGF-like homology #label LE02\
              #domain laminin-type EGF-like homology #label LE03\
              #domain laminin-type EGF-like homology #label LE04\
              #domain laminin-type EGF-like homology #label LE05\
              #domain IV #label DOM4\
              #domain laminin-type EGF-like homology #label LE06\
              #domain III #label DOM3\
              #domain laminin-type EGF-like homology #label LE07\
              #domain laminin-type EGF-like homology #label LE08\
              #domain laminin-type EGF-like homology #label LE09\
              #domain laminin-type EGF-like homology #label LE10\
              #domain laminin-type EGF-like homology #label LE11\
              #domain laminin-type EGF-like homology #label LE12\
              #domain laminin-type EGF-like homology #label LE13\
              #domain II, heptad repeats #label DOM2\
              #domain alpha #label ALP\
              #domain I, heptad repeats #label DOM1\
              #disulfide_bonds #status predicted\
              #binding_site carbohydrate (Asn) (covalent) #status
              predicted\
              #disulfide_bonds interchain #status predicted
              #length 1801 #molecular-weight 196472 #checksum 8852
SUMMARY
Query Match   72.9%; Score 43; DB 1; Length 1801;
Best Local Similarity 50.0%; Pred. No. 1.57e+01;
Matches 1; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 890 DAHTGACL 897
|||||
QY 78 EGTSGCL 85

RESULT 11
ENTRY   C65065 #type complete
TITLE   preprolactin peptidase dependent protein A precursor -
              Escherichia coli (strain K-12)

```

```

ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
ACCESSIONS C65065
REFERENCE A64720
#authors
Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession C65065
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-156 #label BLAT
#cross-references GB:AE000366; GB:U00096; NID:g1789190;
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene
SUMMARY
#length 156 #molecular-weight 17865 #checksum 5159
Query Match 71.2%; Score 42; DB 2; Length 156;
Best Local Similarity 62.5%; Pred. No. 2.56e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 149 QGETEGCL 156
:|:|:|
Qy 78 EGTSGCL 85

RESULT 12
ENTRY #type complete
TITLE hypothetical protein jhp0749 - Helicobacter pylori (strain
J99)
ORGANISM #formal_name Helicobacter pylori
strain J99
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
ACCESSIONS G71891
REFERENCE A71800
#authors
Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
Dolg, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
G.F.; Trust, T.J.
#journal Nature (1999) 397:176-180
#title Genomic sequence comparison of two unrelated isolates of the
human gastric pathogen Helicobacter pylori.
#cross-references MUID:99120557
#accession G71891
#status preliminary
#molecule_type DNA
#residues 1-205 #label ARN
#cross-references GB:AE001506; GB:AE001439; NID:g4155310;
PIDN:AAD06340.1; PID:g4155325
#experimental_source strain J99
GENETICS
#gene jhp0749
CLASSIFICATION #superfamily glyoxalase
SUMMARY #length 205 #molecular-weight 23367 #checksum 846
Query Match 71.2%; Score 42; DB 2; Length 205;
Best Local Similarity 83.3%; Pred. No. 2.56e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 124 GHTPGC 129
:|:|:|
Qy 79 GHTSGC 84

ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
ACCESSIONS C65065
REFERENCE A64720
#authors
Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession C65065
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-156 #label BLAT
#cross-references GB:AE000366; GB:U00096; NID:g1789190;
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene
SUMMARY
#length 156 #molecular-weight 17865 #checksum 5159
Query Match 71.2%; Score 42; DB 2; Length 156;
Best Local Similarity 62.5%; Pred. No. 2.56e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 149 QGETEGCL 156
:|:|:|
Qy 78 EGTSGCL 85

RESULT 12
ENTRY #type complete
TITLE hypothetical protein jhp0749 - Helicobacter pylori (strain
J99)
ORGANISM #formal_name Helicobacter pylori
strain J99
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
ACCESSIONS G71891
REFERENCE A71800
#authors
Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
Dolg, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
G.F.; Trust, T.J.
#journal Nature (1999) 397:176-180
#title Genomic sequence comparison of two unrelated isolates of the
human gastric pathogen Helicobacter pylori.
#cross-references MUID:99120557
#accession G71891
#status preliminary
#molecule_type DNA
#residues 1-205 #label ARN
#cross-references GB:AE001506; GB:AE001439; NID:g4155310;
PIDN:AAD06340.1; PID:g4155325
#experimental_source strain J99
GENETICS
#gene jhp0749
CLASSIFICATION #superfamily glyoxalase
SUMMARY #length 205 #molecular-weight 23367 #checksum 846
Query Match 71.2%; Score 42; DB 2; Length 205;
Best Local Similarity 83.3%; Pred. No. 2.56e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 124 GHTPGC 129
:|:|:|
Qy 79 GHTSGC 84

ORGANISM #formal_name Helicobacter pylori
DATE 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
ACCESSIONS E64621
REFERENCE A64520
#authors
Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
Richardson, D.; Dodson, R.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
McKenny, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
C.M.; Venter, J.C.
#journal Nature (1997) 388:539-547
#title The complete genome sequence of the gastric pathogen
Helicobacter pylori.
#cross-references MUID:97394467
#accession E64621
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-205 #label TOM
#cross-references GB:AE000593; GB:AE000511; NID:g2313944;
PIDN:AAD07862.1; PID:g2313945; TIGR:HP0813
GENETICS
#start_codon TTG
CLASSIFICATION #superfamily glyoxalase
SUMMARY #length 205 #molecular-weight 23463 #checksum 441
Query Match 71.2%; Score 42; DB 2; Length 205;
Best Local Similarity 83.3%; Pred. No. 2.56e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 124 GHTPGC 129
:|:|:|
Qy 79 GHTSGC 84

RESULT 14
ENTRY #type complete
TITLE glutamate--ammonia ligase (EC 6.3.1.2) II - Bradyrhizobium
japonicum
ALTERNATE_NAMES glutamine synthetase II
ORGANISM #formal_name Bradyrhizobium japonicum
DATE 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change
ACCESSIONS A24155
REFERENCE A24155
#authors
Carlson, T.A.; Chelm, B.K.
#journal Nature (1986) 322:568-570
#title Apparent eukaryotic origin of glutamine synthetase II from
the bacterium Bradyrhizobium japonicum.
#accession A24155
#molecule_type DNA
#residues 1-329 #label CAR
COMMENT Two forms of glutamine synthetase (GSI and GSII) can be found in B.
japonicum, a soybean symbiont that belongs to the Rhizobiaceae
family of bacteria. GSI is a typical prokaryotic glutamine
synthetase whereas GSII is similar to the eukaryotic enzyme.
GENETICS
#gene glnII
CLASSIFICATION #superfamily glutamate--ammonia ligase
KEYWORDS ligase
SUMMARY #length 329 #molecular-weight 36906 #checksum 8065

```


Query Match 71.2%; Score 42; DB 1; Length 329;
 Best Local Similarity 62.5%; Pred. No. 2.56e+01;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 52 EGHSSDCV 59
 |||:|:
 QY 78 EGTSGCL 85

RESULT 15
 ENTRY AJSM2V #type complete
 TITLE glutamate--ammonia ligase (EC 6.3.1.2) II - Streptomyces
 viridochromogenes
 ALTERNATE_NAMES glutamine synthetase II
 ORGANISM #formal_name Streptomyces viridochromogenes
 DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
 18-Jun-1999
 ACCESSIONS C36724; S09195; S11635
 REFERENCE A36724
 #authors Behrmann, I.; Hillemann, D.; Puehler, A.; Strauch, E.;
 Wohlleben, W.
 #journal J. Bacteriol. (1990) 172:5326-5334
 #title Overexpression of a Streptomyces viridochromogenes gene
 (glnII) encoding a glutamine synthetase similar to those of
 eucaryotes confers resistance against the antibiotic
 phosphinothricyl-alanyl-alanine.
 #cross-references MUID:90368592
 #accession C36724
 #molecule_type DNA
 #residues 1-343 #label BEH
 ##cross-references EMBL:X52842; NID:g47983; PIDN:CAA37028.1; PID:g581783

GENETICS
 #gene glnII
 #start_codon GTG
 CLASSIFICATION #superfamily glutamate--ammonia ligase
 KEYWORDS ligase
 SUMMARY #length 343 #molecular_weight 37258 #checksum 4274

Query Match 71.2%; Score 42; DB 1; Length 343;
 Best Local Similarity 62.5%; Pred. No. 2.56e+01;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 50 EGHSSDCV 57
 |||:|:
 QY 78 EGTSGCL 85

Search completed: Wed May 10 12:42:12 2000
 Job time : 9 secs.

MORF

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 12:35:53 2000; MasPar time 88.56 Seconds
Tabular output not generated. 2.751 Million cell updates/sec

Title: >US-09-376-430-2
Description: (78-85) from US09376430A.pep (12 of 25)
Perfect Score: 59
Sequence: 1 EIGHTSGCL 8

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swissprot
1:swissprot

Statistics: Mean 20.428; Variance 21.559; scale 0.948

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	46	78.0	503	1	YB66_SCHPO PUTATIVE ATP-DEPENDENT	1.13e+00
2	45	76.3	1959	1	AGRI_RAT AGRIN PRECURSOR.	1.98e+00
3	44	74.6	291	1	LCAT_CLEGL PHOSPHATIDYLCHOLINE-ST	3.43e+00
4	44	74.6	1955	1	AGRI_CHICK AGRIN PRECURSOR.	3.43e+00
5	43	72.9	902	1	ANDR_RAT ANDROGEN RECEPTOR.	5.90e+00
6	43	72.9	1207	1	DPOL_ASFB7 DNA POLYMERASE (EC 2.7	5.90e+00
7	43	72.9	1244	1	DPOL_ASFL6 DNA POLYMERASE (EC 2.7	5.90e+00
8	43	72.9	1801	1	LMB2_RAT LAMININ BETA-2 CHAIN P	5.90e+00
9	42	71.2	156	1	PPDA_ECOLI PREPELIN PEPTIDASE DEP	1.00e+01
10	42	71.2	312	1	DBXK_ANAPL PUTATIVE STEROID DEHYD	1.00e+01
11	42	71.2	329	1	GLN2_BRAJA GLUTAMINE SYNTHETASE I	1.00e+01
12	42	71.2	343	1	GLN2_STRVR GLUTAMINE SYNTHETASE I	1.00e+01
13	42	71.2	609	1	FETA_HUMAN ALPHA-FETOPROTEIN PREC	1.00e+01
14	41	69.5	438	1	LCAT_MOUSE PHOSPHATIDYLCHOLINE-ST	1.70e+01
15	41	69.5	440	1	LCAT_RABIT PHOSPHATIDYLCHOLINE-ST	1.70e+01
16	41	69.5	440	1	LCAT_RAT PHOSPHATIDYLCHOLINE-ST	1.70e+01
17	41	69.5	440	1	LCAT_HUMAN PHOSPHATIDYLCHOLINE-ST	1.70e+01
18	41	69.5	440	1	LCAT_PAPAN PHOSPHATIDYLCHOLINE-ST	1.70e+01
19	41	69.5	480	1	BIND_ARBPB BINDIN PRECURSOR.	1.70e+01
20	41	69.5	545	1	HYAL_RABIT HYALURONIDASE PRECURSOR	1.70e+01
21	41	69.5	1639	1	LMGL_DROME LAMININ GAMMA-1 CHAIN	1.70e+01
22	41	69.5	1758	1	CA24_CAEEL COLLAGEN ALPHA 2(IV) C	1.70e+01
23	41	69.5	1763	1	CA24_ASCSU COLLAGEN ALPHA 2(IV) C	1.70e+01

24	24	41	69.5	1799	1	LMB2_MOUSE	LAMININ BETA-2 CHAIN P	1.70e+01
25	41	69.5	2344	1	POLN_RHD	NON-STRUCTURAL POLYPRO	1.70e+01	
26	40	67.8	119	1	US04_HCMVA	HYPOTHETICAL PROTEIN U	2.83e+01	
27	40	67.8	190	1	SLYD_HAEIN	FKBP-TYPE PEPTIDYL-PRO	2.83e+01	
28	40	67.8	194	1	CLP_HUMAN	LIM DOMAIN PROTEIN, CA	2.83e+01	
29	40	67.8	194	1	CLP_RAT	LIM DOMAIN PROTEIN, CA	2.83e+01	
30	40	67.8	194	1	CLP_MOUSE	LIM DOMAIN PROTEIN, CA	2.83e+01	
31	40	67.8	224	1	GPH_HAEIN	PHOSPHOGLYCOLATE PHOSP	2.83e+01	
32	40	67.8	446	1	GLNA_METPO	GLUTAMINE SYNTHETASE (2.83e+01	
33	40	67.8	446	1	GLNA_METMP	GLUTAMINE SYNTHETASE (2.83e+01	
34	40	67.8	456	1	YD45_SCHPO	HYPOTHETICAL 50.5 KD P	2.83e+01	
35	40	67.8	459	1	LDL1_SCHPO	DIHYDROLIPOAMIDE DEHYD	2.83e+01	
36	40	67.8	584	1	ENV1_HUMAN	RETROVIRUS-RELATED ENV	2.83e+01	
37	40	67.8	609	1	FETA_GORGO	ALPHA-FETOPROTEIN PREC	2.83e+01	
38	40	67.8	1020	1	GCSP_HUMAN	GLYCINE-FETOPROTEINASE	2.83e+01	
39	40	67.8	1147	1	TEAL_SCHPO	TIP ELONGATION ABERRAN	2.83e+01	
40	40	67.8	1387	1	RGSC_RAT	REGULATOR OF G-PROTEIN	2.83e+01	
41	40	67.8	1789	1	LMB1_DROME	LAMININ BETA-1 CHAIN P	2.83e+01	
42	39	66.1	261	1	PRC9_RAT	PROTEASOME COMPONENT C	4.69e+01	
43	39	66.1	448	1	FDF1_CANAL	FARNESYL-DIPHOSPHATE F	4.69e+01	
44	39	66.1	454	1	RADA_CHLTR	RNA REPAIR PROTEIN RAD	4.69e+01	
45	39	66.1	861	1	ORC1_HUMAN	ORIGIN RECOGNITION COM	4.69e+01	

ALIGNMENTS

RESULT	1
ID	YB66_SCHPO
AC	Q09747;
DT	01-NOV-1995 (Rel. 32, Created)
DT	01-NOV-1995 (Rel. 32, Last sequence update)
DT	15-JUL-1999 (Rel. 38, Last annotation update)
DE	PUTATIVE ATP-DEPENDENT RNA HELICASE C12C2.06.
GN	SPBC12C2.06.
OS	Schizosaccharomyces pombe (Fission yeast).
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC	Schizosaccharomycetaceae; Schizosaccharomycetes.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-972;
RA	Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL	Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC	FUNCTION: PUTATIVE ATP-DEPENDENT RNA HELICASE.
CC	SIMILARITY: TO OTHER "DEAD" BOX FAMILY HELICASES. BELONGS TO THE
CC	EIF4A SUBFAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; Z54140; CAA90819.1; .
DR	PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
DR	PFAM; PF00270; DEAD; 1.
DR	PFAM; PF00271; helicase_C; 1.
KW	Hypothetical protein; ATP-binding; RNA-binding; Helicase.
FT	NP_BIND 160 167
FT	SITE 261 264
SQ	SEQUENCE 503 AA; 56305 MW; 68C7C88D6AED252F CRC64;

Query Match: 78.0%; Score 46; DB 1; Length 503;
Best Local Similarity: 62.5%; Pred. No. 1.13e+00;
Matches: 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 375 DGH7VACL 382
Qy 78 EIGHTSGCL 85

RESULT 2

ID AC AGRI_RAT STANDARD; PRT; 1959 AA.
 P25304; Q63034;
 01-MAY-1992 (Rel. 22, Created)
 01-JUL-1993 (Rel. 26, Last sequence update)
 15-FEB-2000 (Rel. 39, Last annotation update)
 DE AGRIN PRECURSOR.
 GN AGRN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
 RC TISSUE-EMBRYONIC SPINAL CORD;
 RX MEDLINE: 91222570.
 RA Rupp F., Payan D.G., Magill-Solc C., Cowan D.M., Scheller R.H.;
 RT "Structure and expression of a rat agrin.";
 RL Neuron 6:811-823(1991).
 RN [2]
 RP SEQUENCE OF 1777-1801 FROM N.A.
 RX MEDLINE: 92407628.
 RA Rupp F., Oezcelik T., Linial M., Peterson K., Francke U., Scheller R.;
 RT "Structure and chromosomal localization of the mammalian agrin gene.";
 RL J. Neurosci. 12:3535-3544(1992).
 CC -!- FUNCTION: COMPONENT OF THE BASAL LAMINA THAT CAUSES THE
 CC AGGREGATION OF ACETYLCHOLINE RECEPTORS AND ACETYLCHOLINE-ESTERASE
 CC ON THE SURFACE OF MUSCLE FIBERS OF THE NEUROMUSCULAR JUNCTION.
 CC -!- SUBCELLULAR LOCATION: SYNAPTIC BASAL LAMINA AT THE NEUROMUSCULAR
 CC JUNCTION.
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST FIVE DIFFERENT FORMS ARISE BY
 CC ALTERNATIVE SPLICING, THEY DIFFER IN THEIR ACETYLCHOLINE RECEPTOR
 CC CLUSTERING ACTIVITY.
 CC -!- TISSUE SPECIFICITY: EMBRYONIC NERVOUS SYSTEM AND MUSCLE.
 CC -!- DEVELOPMENTAL STAGE: MORE ABUNDANT EARLY IN DEVELOPMENT.
 CC -!- SIMILARITY: CONTAINS 8 KAZAL-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
 CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-18 OR MET-24 IS THE
 CC INITIATOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M64780; AAA40703.1; -.
 DR EMBL; M64780; AAA40702.1; ALT_INIT.
 DR EMBL; S44194; AAB23326.1; -.
 DR FIR; JH0399; AGRT.
 DR HSP; P00740; IIXA.
 DR PRINTS; P00290; KAZALINHBTR.
 DR PROSITE; PS00022; EGF_1; 6.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
 DR PFAM; PF00050; Kazal; 9.
 DR PFAM; PF00053; laminin_EGF; 2.
 DR PFAM; PF00054; laminin_G; 3.
 DR PFAM; PF01390; SEA; 1.
 KW Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
 KW Laminin EGF-like domain.
 FT SIGNAL 1 29
 FT CHAIN 30 1959
 FT DOMAIN 65 137
 FT KAZAL-LIKE 1.
 FT DOMAIN 141 212
 FT KAZAL-LIKE 2.
 FT DOMAIN 213 284
 FT KAZAL-LIKE 3.
 FT DOMAIN 287 356
 FT KAZAL-LIKE 4.
 FT DOMAIN 361 429
 FT KAZAL-LIKE 5.
 FT DOMAIN 430 494
 FT KAZAL-LIKE 6.
 FT DOMAIN 495 559
 FT KAZAL-LIKE 7.
 FT DOMAIN 563 645
 FT KAZAL-LIKE 8.
 FT DOMAIN 688 741
 FT LAMININ EGF-LIKE 1.

FT DOMAIN	742	788	LAMININ EGF-LIKE 2.
FT DOMAIN	794	864	KAZAL-LIKE.
FT DOMAIN	1220	1258	EGF-LIKE 1.
FT DOMAIN	1440	1477	EGF-LIKE 2.
FT DOMAIN	1479	1516	EGF-LIKE 3.
FT DOMAIN	1709	1748	EGF-LIKE 4.
FT DOMAIN	869	992	SER/THR-RICH.
FT DOMAIN	1147	1215	SER/THR-RICH.
FT DISULFID	97	116	POTENTIAL.
FT DISULFID	105	137	POTENTIAL.
FT DISULFID	171	191	POTENTIAL.
FT DISULFID	180	212	POTENTIAL.
FT DISULFID	244	263	POTENTIAL.
FT DISULFID	252	284	POTENTIAL.
FT DISULFID	316	335	POTENTIAL.
FT DISULFID	324	356	POTENTIAL.
FT DISULFID	389	408	POTENTIAL.
FT DISULFID	397	429	POTENTIAL.
FT DISULFID	454	473	POTENTIAL.
FT DISULFID	462	494	POTENTIAL.
FT DISULFID	518	538	POTENTIAL.
FT DISULFID	527	559	POTENTIAL.
FT DISULFID	604	624	POTENTIAL.
FT DISULFID	613	645	POTENTIAL.
FT DISULFID	688	700	BY SIMILARITY.
FT DISULFID	690	707	BY SIMILARITY.
FT DISULFID	709	718	BY SIMILARITY.
FT DISULFID	721	739	BY SIMILARITY.
FT DISULFID	742	754	BY SIMILARITY.
FT DISULFID	744	761	BY SIMILARITY.
FT DISULFID	763	772	BY SIMILARITY.
FT DISULFID	775	786	BY SIMILARITY.
FT DISULFID	823	843	POTENTIAL.
FT DISULFID	832	864	POTENTIAL.
FT DISULFID	1224	1235	BY SIMILARITY.
FT DISULFID	1229	1246	BY SIMILARITY.
FT DISULFID	1248	1257	BY SIMILARITY.
FT DISULFID	1444	1455	POTENTIAL.
FT DISULFID	1449	1465	POTENTIAL.
FT DISULFID	1467	1476	POTENTIAL.
FT DISULFID	1483	1494	BY SIMILARITY.
FT DISULFID	1488	1504	BY SIMILARITY.
FT DISULFID	1506	1515	BY SIMILARITY.
FT DISULFID	1713	1727	BY SIMILARITY.
FT DISULFID	1721	1736	BY SIMILARITY.
FT DISULFID	1738	1747	BY SIMILARITY.
FT CARBOHYD	145	145	POTENTIAL.
FT CARBOHYD	672	672	POTENTIAL.
FT CARBOHYD	827	827	POTENTIAL.
FT CARBOHYD	957	957	POTENTIAL.
FT VARSPLIC	1144	1152	MISSING (IN ISOFORM 2).
FT VARSPLIC	1780	1798	MISSING (IN ISOFORM 3).
FT VARSPLIC	1788	1798	MISSING (IN ISOFORM 4).
FT VARSPLIC	1780	1787	MISSING (IN ISOFORM 5).
FT VARIANT	314	314	V -> VTCD (IN A VARIANT).
FT SEQUENCE	1959 AA	208645 MW	7FEFDFDAFF89CC31 CRC64;
Query Match	76.3%	Score 45	DB 1; Length 1959;
Best Local Similarity	50.0%	Pred. No. 1.98e+00;	
Matches	4;	Conservative	4; Mismatches 0; Indels 0; Gaps 0;
Db	159 DQQTASCL	166	
QY	78 EIGHTSGCL	85	
RESULT	3		
ID	LCAT_CLEGL	STANDARD;	PRT; 291 AA.
AC	O35502;		
DT	15-FEB-2000 (Rel. 39, Created)		
DT	15-FEB-2000 (Rel. 39, Last sequence update)		
DT	15-FEB-2000 (Rel. 39, Last annotation update)		
DE	PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE (EC 2.3.1.43) (LECITHIN-		

DE CHOLESTEROL ACYLTRANSFERASE) (PHOSPHOLIPID-CHOLESTEROL
DE ACYLTRANSFERASE) (FRAGMENT).
OS LCAT.
GN Clethrionomys glareolus (Bank vole).
ON Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae;
OC Clethrionomys.
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson M., Catzefflis F.M., Briolay J., Mouchiroud D.;
RT "Molecular phylogeny of rodents, with special emphasis on Murids:
RT evidence from nuclear gene LCAT".
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CENTRAL ENZYME IN THE EXTRACELLULAR METABOLISM OF PLASMA
CC LIPOPROTEINS. AMONG OTHER SUBSTRATES IT ESTERIFIES THE FREE
CC CHOLESTEROL TRANSPORTED IN PLASMA LIPOPROTEINS.
CC -!- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + STEROL = STEROL ESTER +
CC 1-ACYLGLYCEROPHOSPHOCHOLINE (PALMITOYL, OLEOYL, AND LINOLEOYL CAN
CC BE TRANSFERRED; A NUMBER OF STEROLS, INCLUDING CHOLESTEROL, CAN
CC ACT AS ACCEPTOR).
CC -!- ENZYME REGULATION: APOLIPOPROTEIN A-I IS A POTENT ACTIVATOR FOR
CC THIS ENZYME.
CC -!- SIMILARITY: PARTIAL WITH LIPASES (PANCREATIC, GASTRIC, HEPATIC,
CC LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; U72300; AAB58990.1; -;
DR EMBL; U72299; AAB58990.1; JOINED.
DR PROSITE; PS00120; LIPASE_SER; 1.
KW Transferase; Acyltransferase; Lipid metabolism; Glycoprotein.
FT NON_TER 1
FT ACT_SITE 125 125 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 291 291
SQ SEQUENCE 291 AA; 35536 MW; B1F7E9D89201219D CRC64;
Query Match 74.6%; Score 44; DB 1; Length 291;
Best Local Similarity 71.4%; Pred. No. 3.43e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 123 GHSVGL 129
QY ||: |||
79 GHTSGL 85
RESULT 4
ID AGRI_CHICK STANDARD; PRT: 1955 AA.
AC P31696;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE AGRIN PRECURSOR.
GN AGRN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=BRAIN;
RX MEDLINE; 92232297.
RA Ts'm K.W.K., Ruegg M.A., Escher G., Kroeger S., McMahon U.J.;
RT "cDNA that encodes active agrin".
RL Neuron 8:677-689(1992).
RN [2]

RP ALTERNATIVE SPLICING.
RX MEDLINE; 92232298.
RA Ruegg M.A., Ts'm K.W.K., Horton S.E., Kroeger S., Escher G.,
RA Gensch E.M., McVahan U.J.;
RT "The agrin gene codes for a family of basal lamina proteins that
RT differ in function and distribution".
RL Neuron 8:691-699(1992).
CC -!- FUNCTION: COMPONENT OF THE BASAL LAMINA THAT CAUSES THE
CC AGGREGATION OF ACETYLCHOLINE RECEPTORS AND ACETYLCHOLINE-ESTERASE
CC ON THE SURFACE OF MUSCLE FIBERS OF THE NEUROMUSCULAR JUNCTION.
CC -!- SUBCELLULAR LOCATION: SYNAPTIC BASAL LAMINA AT THE NEUROMUSCULAR
CC JUNCTION.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST THREE DIFFERENT FORMS ARISE BY
CC ALTERNATIVE SPLICING, THEY DIFFER IN THEIR ACETYLCHOLINE RECEPTOR
CC CLUSTERING ACTIVITY.
CC -!- SIMILARITY: CONTAINS 8 KAZAL-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; M94271; AAA48585.1; -;
DR EMBL; M97371; AAA48586.1; -;
DR EMBL; M97372; -; NOT_ANNOTATED_CDS.
DR PIR; JH0591; AGCH.
DR PRINTS; PR00290; KAZALINHBTR.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR PFAM; PF00008; EGF; 4.
DR PFAM; PF00050; kazal; 9.
DR PFAM; PF00053; laminin_EGF; 2.
DR PFAM; PF00054; laminin_G; 3.
DR PFAM; PF01390; SEA; 1.
KW Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW Laminin EGF-like domain.
FT SIGNAL 1 38
FT CHAIN 39 1955
FT DOMAIN 54 126 KAZAL-LIKE 1.
FT DOMAIN 130 201 KAZAL-LIKE 2.
FT DOMAIN 202 273 KAZAL-LIKE 3.
FT DOMAIN 276 344 KAZAL-LIKE 4.
FT DOMAIN 350 418 KAZAL-LIKE 5.
FT DOMAIN 419 483 KAZAL-LIKE 6.
FT DOMAIN 484 548 KAZAL-LIKE 7.
FT DOMAIN 551 633 KAZAL-LIKE 8.
FT DOMAIN 675 728 LAMININ EGF-LIKE 1.
FT DOMAIN 729 775 LAMININ EGF-LIKE 2.
FT DOMAIN 781 851 KAZAL-LIKE
FT DOMAIN 1229 1265 EGF-LIKE 1.
FT DOMAIN 1446 1483 EGF-LIKE 2.
FT DOMAIN 1485 1522 EGF-LIKE 3.
FT DOMAIN 1714 1752 EGF-LIKE 4.
FT DOMAIN 856 995 SER/THR-RICH.
FT DOMAIN 1150 1219 SER/THR-RICH.
FT DISULFID 86 105 POTENTIAL.
FT DISULFID 94 126 POTENTIAL.
FT DISULFID 160 180 POTENTIAL.
FT DISULFID 169 201 POTENTIAL.
FT DISULFID 233 252 POTENTIAL.
FT DISULFID 241 273 POTENTIAL.
FT DISULFID 304 323 POTENTIAL.
FT DISULFID 312 344 POTENTIAL.
FT DISULFID 378 397 POTENTIAL.
FT DISULFID 386 418 POTENTIAL.
FT DISULFID 443 462 POTENTIAL.

```

FT DISULFID 451 483 POTENTIAL.
FT DISULFID 507 527 POTENTIAL.
FT DISULFID 516 548 POTENTIAL.
FT DISULFID 592 612 POTENTIAL.
FT DISULFID 601 633 POTENTIAL.
FT DISULFID 675 687 BY SIMILARITY.
FT DISULFID 677 694 BY SIMILARITY.
FT DISULFID 696 705 BY SIMILARITY.
FT DISULFID 708 726 BY SIMILARITY.
FT DISULFID 729 741 BY SIMILARITY.
FT DISULFID 731 748 BY SIMILARITY.
FT DISULFID 750 759 BY SIMILARITY.
FT DISULFID 762 773 BY SIMILARITY.
FT DISULFID 810 830 POTENTIAL.
FT DISULFID 819 851 POTENTIAL.
FT DISULFID 1233 1244 BY SIMILARITY.
FT DISULFID 1238 1253 BY SIMILARITY.
FT DISULFID 1255 1264 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1455 1471 BY SIMILARITY.
FT DISULFID 1473 1482 BY SIMILARITY.
FT DISULFID 1489 1500 BY SIMILARITY.
FT DISULFID 1494 1510 BY SIMILARITY.
FT DISULFID 1512 1521 BY SIMILARITY.
FT DISULFID 1718 1731 BY SIMILARITY.
FT DISULFID 1725 1740 BY SIMILARITY.
FT DISULFID 1742 1751 BY SIMILARITY.
FT CARBOHYD 390 390 POTENTIAL.
FT CARBOHYD 659 659 POTENTIAL.
FT CARBOHYD 764 764 POTENTIAL.
FT CARBOHYD 814 814 POTENTIAL.
FT VARSPPLIC 1648 1651 MISSING (IN AGRIN-RELATED PROTEIN 2).
FT VARSPPLIC 1763 1793 AGRIN-RELATED PROTEIN 2).
FT CONFLICT 1129 1131 RII -> SIL (IN AAA48586).
FT SEQUENCE 1955 AA; 211411 MW; B4DEB27C33422581 CRC64;

Query Match 74.6%; Score 44; DB 1; Length 1955;
Best Local Similarity 50.0%; Pred. No. 3.43e+00;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 148 DQOTAGCV 155
Qy 78 EGTSGCL 85

RESULT 5
ID ANDR_RAT STANDARD; PRT; 902 AA.
AC P13207;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ANDROGEN RECEPTOR.
GN AR OR NR3C4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89112209.
RA Tan J., Joseph D.R., Quarby V.E., Lubahn D.B., Sar M., French F.S.,
RA Wilson E.M.;
RT "The rat androgen receptor: primary structure, autoregulation of its
RT messenger ribonucleic acid, and immunocytochemical localization of
RT the receptor protein."
RL Mol. Endocrinol. 2:1276-1285(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-PROSTATE;
RX MEDLINE; 89017168.
RA Chang C., Kokontis J., Liao S.;
RT "Structural analysis of complementary DNA and amino acid sequences of
RT human and rat androgen receptors."

```

```

RL Proc. Natl. Acad. Sci. U.S.A. 85:7211-7215(1988).
CC -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS IN THE SEMINAL VESICLE,
CC VENTRAL PROSTATE AND COAGULATING GLAND WITH LOWER LEVELS IN THE
CC KIDNEY, AND LEVATOR ANI MUSCLE.
CC -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -!- MISCELLANEOUS: IN THE ABSENCE OF LIGAND, STEROID HORMONE RECEPTORS
CC ARE THOUGHT TO BE WEAKLY ASSOCIATED WITH NUCLEAR COMPONENTS;
CC HORMONE BINDING GREATLY INCREASES RECEPTOR AFFINITY. THE
CC HORMONE-RECEPTOR COMPLEX APPEARS TO RECOGNIZE DISCRETE DNA
CC SEQUENCES UPSTREAM OF TRANSCRIPTIONAL START SITES.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M20133; AAA40733.1; -
CC EMBL; M23264; AAA40759.1; -
CC PIR; A34943; A34943.
CC HSP; P06536; IRGD.
CC TRANSFAC; T00042; -
CC PRINTS; PR00047; STROIDFINGER.
CC PRINTS; PR00398; STROHORMONER.
CC PRINTS; PR00521; ANDROGENR.
CC PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
CC PFW; PF00104; hormone_rec; 1.
CC PFW; PF00105; zf-C4; 1.
CC Receptor; Transcription regulation; DNA-binding; Nuclear protein;
CC Zinc-finger; Steroid-binding.
FT DOMAIN 1 540 MODULATING.
FT DNA_BIND 542 607 C4-TYPE ZINC FINGERS (TWO).
FT ZN_FING 542 582 C4-TYPE.
FT ZN_FING 578 602 C4-TYPE.
FT DOMAIN 673 902 LIGAND-BINDING.
FT DOMAIN 63 67 POLY-ARG.
FT DOMAIN 174 195 POLY-GLN.
FT DOMAIN 370 376 POLY-PRO.
FT DOMAIN 394 400 POLY-ALA.
FT DOMAIN 444 450 POLY-GLY.
FT CONFLICT 389 389 S -> L (IN REF. 2).
SQ SEQUENCE 902 AA; 98217 MW; 43F4064759FDCBED CRC64;

Query Match 72.9%; Score 43; DB 1; Length 902;
Best Local Similarity 77.8%; Pred. No. 5.90e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 103 EGHPSGCL 111
Qy 78 EGTSGCL 85

RESULT 6
ID DPOLASFB7 STANDARD; PRT; 1207 AA.
AC P42489;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE (EC 2.7.7.7).
GN DPOL OR G1207R OR G1211R.
OS African swine fever virus (strain BA71V) (ASFV).
OC Viruses; dsDNA viruses, no RNA stage;
OC African swine fever-like viruses.
RN [1]

```

RP SEQUENCE FROM N.A.
 RX MEDLINE: 94123986.
 RA Rodriguez J.M., Yanez R.J., Rodriguez J.F., Vinuela E., Salas M.L.;
 RT "The DNA polymerase-encoding gene of African swine fever virus;
 RL sequence and transcriptional analysis.";
 RN Gene 136:103-110(1993).
 [2]
 RP SEQUENCE FROM N.A.
 RA Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
 RA Rodriguez J.F., Vinuela E.;
 RT "Analysis of the complete nucleotide sequence of African swine fever
 RT virus.";
 RL Virology 208:249-278(1995).
 CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
 CC N PYROPHOSPHATE + DNA(N).
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L19042; AAA42690.1; -;
 CC EMBL: U18466; AAB65319.1; ALT_INIT.
 CC PRINTS: PR00106; DNAPOLB.
 CC PROSITE: PS00116; DNA_POLYMERASE_B; 1.
 CC PFAM: PF00136; DNA_POL_B; 4.
 CC Transferase; DNA-directed DNA polymerase; DNA replication;
 CC DNA-binding; Repeat.
 CC DOMAIN 1072 1087 4 X 4 AA TANDEM REPEATS OF A-G-N-P.
 CC SEQUENCE 1207 AA; 139391 MW; C39F7F7A0F9639DC CRC64;

Query Match 72.9%; Score 43; DB 1; Length 1207;
 Best Local Similarity 83.3%; Pred. No. 5.90e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 353 HTPGCL 358
 ||:||||
 QY 80 HTSGCL 85

RESULT 7
 ID DPOL ASFL6 STANDARD; PRT; 1244 AA.
 AC P43139; Q89937; Q89938; Q89939;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE DNA POLYMERASE (EC 2.7.7.7).
 GN DPOL.
 OS African swine fever virus (isolate Lisbon 60 / Lis60) (ASFV).
 OC Viruses; dsDNA viruses, no RNA stage;
 OC African swine fever-like viruses.
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94167248.
 RA Martins A., Ribeiro G., Marques M.I., Costa J.V.;
 RT "Genetic identification and nucleotide sequence of the DNA polymerase
 RT gene of African swine fever virus.";
 RL Nucleic Acids Res. 22:208-213(1994).
 [2]
 RP SEQUENCE FROM N.A., AND PHOSPHONACETATE-RESISTANT VARIANTS.
 RX MEDLINE: 96095225.
 RA Marques M.I., Costa J.V.;
 RT "Functional and molecular characterization of African swine fever
 RT virus mutants resistant to phosphonacetic acid.";
 RL Virology 214:72-81(1995).
 CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
 CC N PYROPHOSPHATE + DNA(N).
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L19042; AAA42690.1; -;
 CC EMBL: U18466; AAB65319.1; ALT_INIT.
 CC PRINTS: PR00106; DNAPOLB.
 CC PROSITE: PS00116; DNA_POLYMERASE_B; 1.
 CC PFAM: PF00136; DNA_POL_B; 4.
 CC Transferase; DNA-directed DNA polymerase; DNA replication;
 CC DNA-binding; Repeat.
 CC DOMAIN 1072 1087 4 X 4 AA TANDEM REPEATS OF A-G-N-P.
 CC SEQUENCE 1207 AA; 139391 MW; C39F7F7A0F9639DC CRC64;

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X73330; CAA51757.1; -;
 CC EMBL: U27573; AAB60596.1; -;
 CC EMBL: U27574; AAB60597.1; -;
 CC EMBL: U27575; AAB60598.1; -;
 CC PRINTS: PR00106; DNAPOLB.
 CC PROSITE: PS00116; DNA_POLYMERASE_B; 1.
 CC PFAM: PF00136; DNA_POL_B; 4.
 CC Transferase; DNA-directed DNA polymerase; DNA replication;
 CC DNA-binding; Repeat.
 CC DOMAIN 1069 1120 13 X 4 AA TANDEM REPEATS OF A-G-N-P.
 CC VARIANT 436 436 D -> G (IN ISOLATE PAAR-100;
 CC PHOSPHONACETATE-RESISTANT).
 CC VARIANT 512 512 L -> F (IN ISOLATE PAAR-463;
 CC PHOSPHONACETATE-RESISTANT).
 CC VARIANT 515 515 A -> V (IN ISOLATE PAAR-20;
 CC PHOSPHONACETATE-RESISTANT).
 CC SEQUENCE 1244 AA; 142533 MW; 3689D15732B0A19E CRC64;
 Query Match 72.9%; Score 43; DB 1; Length 1244;
 Best Local Similarity 83.3%; Pred. No. 5.90e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 353 HTPGCL 358
 ||:||||
 QY 80 HTSGCL 85
 RESULT 8
 ID LMB2_RAT STANDARD; PRT; 1801 AA.
 AC P15800;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE LAMININ BETA-2 CHAIN PRECURSOR (S-LAMININ) (LAMININ CHAIN B3).
 GN LAMB2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=LIVER;
 RX MEDLINE: 89159410.
 RA Hunter D.D., Shah V., Merlie J.P., Sanes J.R.;
 RT "A laminin-like adhesive protein concentrated in the synaptic cleft
 RT of the neuromuscular junction.";
 RL Nature 338:229-234(1989).
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
 CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC THE BETA-2 CHAIN IS A SUBUNIT OF LAMININ-3 (S-LAMININ), LAMININ-4
 CC (S-MEROSIN), AND LAMININ-7 (KS-LAMININ).
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT). S-LAMININ IS CONCENTRATED IN THE SYNAPTIC
 CC CLEFT OF THE NEUROMUSCULAR JUNCTION.
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.

CC CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC CC the European Bioinformatics Institute. There are no restrictions on its
 CC CC use by non-profit institutions as long as its content is in no way
 CC CC modified and this statement is not removed. Usage by and for commercial
 CC CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC CC or send an email to license@isb-sib.ch).
 CC CC -----
 CC CC EMBL: X16563; CAA34561.1; -
 CC CC DR PIR: S03539; MMRTS.
 CC CC DR HSP: P02458; 1KLO.
 CC CC DR PRINTS: PR00011; EGF-LAMININ.
 CC CC DR PROSITE: PS00022; EGF_1; 10.
 CC CC DR PROSITE: PS01186; EGF_2; 2.
 CC CC DR PROSITE: PS01248; LAMININ_TYPE_EGF; 12.
 CC CC DR PFAM: PF00053; laminin_EGF; 13.
 CC CC DR PFAM: PF00055; laminin_Nterm; 1.
 CC CC KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 CC KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 CC FT SIGNAL 1 35
 CC FT CHAIN 36 1801 LAMININ BETA-2 CHAIN.
 CC FT DOMAIN 36 283 LAMININ N-TERMINAL (DOMAIN VI). (DOMAIN
 CC FT DOMAIN 286 555 4.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
 CC FT V).
 CC FT DOMAIN 286 349 LAMININ EGF-LIKE 1.
 CC FT DOMAIN 350 412 LAMININ EGF-LIKE 2.
 CC FT DOMAIN 413 472 LAMININ EGF-LIKE 3.
 CC FT DOMAIN 473 524 LAMININ EGF-LIKE 4.
 CC FT DOMAIN 525 555 LAMININ EGF-LIKE 5. (INCOMPLETE).
 CC FT DOMAIN 556 785 LAMININ EGF-LIKE 6.
 CC FT DOMAIN 786 1192 8 X LAMININ EGF-LIKE REPEATS (DOMAIN
 CC FT III).
 CC FT DOMAIN 786 833 LAMININ EGF-LIKE 6.
 CC FT DOMAIN 834 879 LAMININ EGF-LIKE 7.
 CC FT DOMAIN 880 929 LAMININ EGF-LIKE 8.
 CC FT DOMAIN 930 988 LAMININ EGF-LIKE 9.
 CC FT DOMAIN 989 1040 LAMININ EGF-LIKE 10.
 CC FT DOMAIN 1041 1097 LAMININ EGF-LIKE 11.
 CC FT DOMAIN 1098 1145 LAMININ EGF-LIKE 12.
 CC FT DOMAIN 1146 1192 LAMININ EGF-LIKE 13.
 CC FT DOMAIN 1193 1412 DOMAIN II.
 CC FT DOMAIN 1413 1445 DOMAIN ALPHA.
 CC FT DOMAIN 1446 1801 DOMAIN I.
 CC FT DOMAIN 1259 1306 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1475 1529 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1576 1793 COILED COIL (POTENTIAL).
 CC FT DOMAIN 286 295 BY SIMILARITY.
 CC FT DISULFID 288 313 BY SIMILARITY.
 CC FT DISULFID 315 324 BY SIMILARITY.
 CC FT DISULFID 327 347 BY SIMILARITY.
 CC FT DISULFID 350 359 BY SIMILARITY.
 CC FT DISULFID 352 377 BY SIMILARITY.
 CC FT DISULFID 380 389 BY SIMILARITY.
 CC FT DISULFID 392 410 BY SIMILARITY.
 CC FT DISULFID 413 426 BY SIMILARITY.
 CC FT DISULFID 441 441 BY SIMILARITY.
 CC FT DISULFID 443 452 BY SIMILARITY.
 CC FT DISULFID 455 470 BY SIMILARITY.
 CC FT DISULFID 473 487 BY SIMILARITY.
 CC FT DISULFID 494 494 BY SIMILARITY.
 CC FT DISULFID 496 505 BY SIMILARITY.
 CC FT DISULFID 508 522 BY SIMILARITY.
 CC FT DISULFID 786 798 BY SIMILARITY.
 CC FT DISULFID 788 805 BY SIMILARITY.
 CC FT DISULFID 807 816 BY SIMILARITY.
 CC FT DISULFID 819 831 BY SIMILARITY.
 CC FT DISULFID 834 846 BY SIMILARITY.
 CC FT DISULFID 856 853 BY SIMILARITY.
 CC FT DISULFID 855 864 BY SIMILARITY.
 CC FT DISULFID 867 877 BY SIMILARITY.
 CC FT DISULFID 880 889 BY SIMILARITY.

FT DISULFID 882 896 BY SIMILARITY.
 FT DISULFID 899 908 BY SIMILARITY.
 FT DISULFID 911 927 BY SIMILARITY.
 FT DISULFID 930 946 BY SIMILARITY.
 FT DISULFID 932 957 BY SIMILARITY.
 FT DISULFID 959 968 BY SIMILARITY.
 FT DISULFID 971 986 BY SIMILARITY.
 FT DISULFID 989 1003 BY SIMILARITY.
 FT DISULFID 991 1010 BY SIMILARITY.
 FT DISULFID 1013 1022 BY SIMILARITY.
 FT DISULFID 1025 1038 BY SIMILARITY.
 FT DISULFID 1098 1110 BY SIMILARITY.
 FT DISULFID 1100 1117 BY SIMILARITY.
 FT DISULFID 1119 1128 BY SIMILARITY.
 FT DISULFID 1131 1143 BY SIMILARITY.
 FT DISULFID 1146 1158 BY SIMILARITY.
 FT DISULFID 1148 1165 BY SIMILARITY.
 FT DISULFID 1167 1176 BY SIMILARITY.
 FT DISULFID 1179 1190 BY SIMILARITY.
 FT DISULFID 1193 1193 INTERCHAIN (PROBABLE).
 FT DISULFID 1196 1196 INTERCHAIN (PROBABLE).
 FT DISULFID 1800 1800 INTERCHAIN (PROBABLE).
 FT CARBOHYD 251 251 POTENTIAL.
 FT CARBOHYD 371 371 POTENTIAL.
 FT CARBOHYD 1083 1088 POTENTIAL.
 FT CARBOHYD 1252 1252 POTENTIAL.
 FT CARBOHYD 1311 1311 POTENTIAL.
 FT CARBOHYD 1351 1351 POTENTIAL.
 FT CARBOHYD 1502 1502 POTENTIAL.
 SQ SEQUENCE 1801 AA; 196473 MW; 97AEF32F8F31FA75 CRC64;
 Query Match 72.9%; Score 43; DB 1; Length 1801;
 Best Local Similarity 50.0%; Pred. No. 5.90e+00;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 890 DAHTGACL 897
 Qy 78 EGH7SGCL 85
 RESULT 9
 ID PFDA_ECOLI STANDARD; PRT; 156 AA.
 AC P33554; Q59379;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PREPILIN PEPTIDASE DEPENDENT PROTEIN A PRECURSOR.
 GN PFDA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 86232583.
 RA Finch P.W., Willson R.E., Brown K., Hickson I.D., Tomkinson A.E.,
 RA Emerson P.T.;
 RA "Complete nucleotide sequence of the Escherichia coli recC gene and
 RA of the thyA-recC intergenic region.";
 RA Nucleic Acids Res. 14:4437-4451(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE; 97426617.
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Gollado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RA "The complete genome sequence of Escherichia coli K-12.";
 RA Science 277:1453-1474(1997).
 RN [3]
 RP IDENTIFICATION.
 RX MEDLINE; 95020523.
 RA Hobbs M., Mattick J.S.;

RT "Common components in the assembly of type 4 fimbriae, DNA transfer
RT systems, filamentous phage and protein-secretion apparatus: a general
RT system for the formation of surface-associated protein complexes.";
RL Mol. Microbiol. 10:233-243(1993).
CC -1- FUNCTION: NOT YET KNOWN.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X03966; -; NOT_ANNOTATED_CDS.
DR EMBL; U29581; AAB40473.1; -;
DR EMBL; AE000366; AAC75865.1; -;
DR EGOGENE; EG12081; PFDA.
DR PROSITE; PS00409; PROKAR_NTER_METHYL; 1.
KW Methylation.
FT PROPEP 1 6 BY SIMILARITY.
FT CHAIN 7 156 PREPILIN PEPTIDASE DEPENDENT PROTEIN A.
FT MOD_RES 7 7 METHYLATION (BY SIMILARITY).
FT CONFLICT 5 5 A -> R (IN REF. 2).
SQ SEQUENCE 156 AA; 17780 MW; 3D32A17839EB951C CRC64;

Query Match 71.2%; Score 42; DB 1; Length 156;
Best Local Similarity 62.5%; Pred. No. 1.00e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 149 QGETEGCL 156

:| | | |
QY 78 EGHSTGCL 85

RESULT 10 STANDARD; PRT; 312 AA.
ID DBXNANPL
AC OS7314;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PUTATIVE STEROID DEHYDROGENASE SPM2 (EC 1.1.1.1.-).
OS Anas platyrhynchos (domestic duck).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Anseriformes; Anatidae; Anas.
[1]
RP SEQUENCE FROM N.A.
RC TISSUE-FOREBRAIN;
RA Kimura N., Kurosawa N., Kondo K., Tsukada Y.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY. 17-BETA-HSD 3 SUBFAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AB009304; BAA23765.1; -;
DR PRINTS; PR00080; ALCDHDRGNASE.
DR PROSITE; PS00061; ADH_SHORT; 1.
DR PFAM; PF00106; adh_short; 1.
KW Steroid biosynthesis; Oxidoreductase; NADP; Multigene family.
FT NP_BIND 48 77 NADP (BY SIMILARITY).
FT ACT_SITE 201 201 BY SIMILARITY.
SQ SEQUENCE 312 AA; 33896 MW; 744D8F4491A9262 CRC64;

Query Match 71.2%; Score 42; DB 1; Length 312;
Best Local Similarity 50.0%; Pred. No. 1.00e+01;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 265 QSTNGCL 272
:| | | |
QY 78 EGHSTGCL 85

RESULT 11 STANDARD; PRT; 329 AA.
ID GLN2_BRAJA
AC P04772;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUTAMINE SYNTHETASE II (EC 6.3.1.2) (GLUTAMATE--AMMONIA LIGASE II)
DE (GSII).
DE (GSII).
GN GLNII.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Bradyrhizobium group;
OC Bradyrhizobium.
CC [1]
RP SEQUENCE FROM N.A.
RA Carlson T.A., Chelm B.K.;
RT "Apparent eukaryotic origin of glutamine synthetase II from the
RT bacterium Bradyrhizobium japonicum.";
RL Nature 322:568-570(1986).
CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) = ADP + GLUTAMINE +
CC ORTHOPHOSPHATE.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- MISCELLANEOUS: TWO FORMS OF GLUTAMINE SYNTHETASE (GSI AND GSII)
CC CAN BE FOUND IN THIS NITROGEN FIXING BACTERIA, GSI IS A TYPICAL
CC PROKARYOTIC GLUTAMINE SYNTHETASE WHEREAS GSII IS SIMILAR TO THE
CC EUKARYOTIC ENZYME.
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X04187; CAA27779.1; -;
DR PIR; A24155; AJZJ02.
DR PROSITE; PS00180; GLNA_1; 1.
DR PROSITE; PS00181; GLNA_ATP; 1.
DR PFAM; PF00120; gln-synt; 1.
KW Nitrogen fixation; Ligase; Multigene family.
SQ SEQUENCE 329 AA; 36906 MW; 3EAF21D715F9165 CRC64;

Query Match 71.2%; Score 42; DB 1; Length 329;
Best Local Similarity 62.5%; Pred. No. 1.00e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 52 EGHSSDCV 59
:| | | |
QY 78 EGHSTGCL 85

RESULT 12 STANDARD; PRT; 343 AA.
ID GLN2_STRVR STANDARD;
AC P19432;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUTAMINE SYNTHETASE II (EC 6.3.1.2) (GLUTAMATE--AMMONIA LIGASE II)
DE (GSII).
DE (GSII).
GN GLNII.
OS Streptomyces viridochromogenes.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomyces.
CC [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN-ES2;
RX MEDLINE; 90368592.
RA Behrmann I., Hillemann D., Puehler A., Strauch E., Wohlleben W.;
RT "Overexpression of a Streptomyces viridochromogenes gene (glnII)
RT encoding a glutamine synthetase similar to those of eucaryotes
RT confers resistance against the antibiotic
RT phosphinothricyl-alanyl-alanine."
RL J. Bacteriol. 172:5326-5334(1990).
CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) -> ADP + GLUTAMINE +
CC ORTHOPHOSPHATE.
CC -!- ENZYME REGULATION: THE ACTIVITY OF THIS ENZYME IS NOT CONTROLLED
CC BY ADENYLATION.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- MISCELLANEOUS: OVEREXPRESSION OF STREPTOMYCES VIRIDCHROMOGENES
CC GLUTAMINE SYNTHETASE CONFERS RESISTANCE AGAINST THE ANTIBIOTIC
CC PHOSPHINOTHRICYL-ALANYL-ALANINE.
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52842; CAA37028.1;
CC PIR; C36724; AJSW2V.
CC PIR; S09195; S09195.
CC PROSITE; PS00180; GLNA_1; 1.
CC PROSITE; PS00181; GLNA_ATP; 1.
CC PFAM; PF00120; gln-synt; 1.
CC Lgase; Multigene family.
CC SEQUENCE 343 AA; 37258 MW; 537C8F425E2E660F CRC64;
CC -----
Query Match 71.2%; Score 42; DB 1; Length 343;
Best Local Similarity 62.5%; Pred. No. 1.00e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 50 EGHSSDCV 57
Qy 78 EGHSSGCL 85
|||||
DEFEFOPROTEIN.

RESULT 13
ID FETA_HUMAN STANDARD; PRT; 509 AA.
AC P02771;
DC 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE ALPHA-FETOPROTEIN PRECURSOR (ALPHA-FETOglobulin) (ALPHA-1-
DE FETOPROTEIN).
GN AFP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 83273664.
RA Morinaga T., Sakai M., Wegmann T.G., Tamaoki T.;
RT "Primary structures of human alpha-fetoprotein and its mRNA."
RL Proc. Natl. Acad. Sci. U.S.A. 80:4604-4608(1983).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87185436.
RA Gibbs P.E.M., Zielinski R., Boyd C., Dugaiczky A.;
RT "Structure, polymorphism, and novel repeated DNA elements revealed by
RT a complete sequence of the human alpha-fetoprotein gene."
RL Biochemistry 26:1332-1343(1987).
RN [3]
RP SEQUENCE OF 1-28 FROM N.A.

```

```

RX MEDLINE; 93278385.
RA McVey J.H., Michaelides K., Hansen L.P., Ferguson-Smith M.,
RA Tilghman S., Krumlauf R., Tuddenham E.G.D.;
RT "A G->A substitution in an HNF 1 binding site in the human alpha-
RT fetoprotein gene is associated with hereditary persistence of alpha-
RT fetoprotein (HPAFP).";
RL Hum. Mol. Genet. 2:379-379(1993).
RN [4]
RP SEQUENCE OF 429-556 FROM N.A.
RX MEDLINE; 83158778.
RA Beattie W.G., Dugaiczky A.;
RT "Structure and evolution of human alpha-fetoprotein deduced from
RT partial sequence of cloned cDNA."
RL Gene 20:415-422(1982).
RN [5]
RP PARTIAL SEQUENCE OF 19-609.
RX MEDLINE; 91242409.
RA Pucci P., Siciliano R., Malorni A., Marino G., Tecce M.F.,
RA Ceccarini C., Terrana B.;
RT "Human alpha-fetoprotein primary structure: a mass spectrometric
RT study."
RL Biochemistry 30:5061-5066(1991).
RN [6]
RP PRELIMINARY SEQUENCE OF 19-35.
RX MEDLINE; 77242506.
RA Yachnin S., Hsu R., Heinrichson R.L., Miller J.B.;
RT "Studies on human alpha-fetoprotein. Isolation and characterization
RT of monomeric and polymeric forms and amino-terminal sequence
RT analysis."
RL Biochim. Biophys. Acta 493:418-428(1977).
RN [7]
RP PRELIMINARY SEQUENCE OF 19-38.
RX MEDLINE; 78001760.
RA Aoyagi Y., Ikenaka T., Ichida F.;
RT "Comparative chemical structures of human alpha-fetoproteins from
RT fetal serum and from ascites fluid of a patient with hepatoma."
RL Cancer Res. 37:3663-3667(1977).
RN [8]
RP PRELIMINARY SEQUENCE OF 19-39.
RX MEDLINE; 75018719.
RA Ruoslahti E., Pihko H., Vaheeri A., Seppala M., Virolainen M.,
RA Kontinen A.;
RT "Alpha fetoprotein: structure and expression in man and inbred mouse
RT strains under normal conditions and liver injury."
RL Johns Hopkins Med. J. Suppl. 3:249-255(1974).
RN [9]
RP GENE STRUCTURE.
RX MEDLINE; 85182629.
RA Sakai M., Morinaga T., Urano Y., Watanabe K., Wegmann T.G.,
RA Tamaoki T.;
RT "The human alpha-fetoprotein gene. Sequence organization and the 5'
RT flanking region."
RL J. Biol. Chem. 260:5055-5060(1985).
RN [10]
RP METAL-BINDING.
RX MEDLINE; 79001617.
RA Aoyagi Y., Ikenaka T., Ichida F.;
RT "Copper(II)-binding ability of human alpha-fetoprotein."
RL Cancer Res. 38:3483-3486(1978).
RN [11]
RP BILIRUBIN BINDING.
RX MEDLINE; 80001710.
RA Aoyagi Y., Ikenaka T., Ichida F.;
RT "Alpha-Fetoprotein as a carrier protein in plasma and its bilirubin-
RT binding ability."
RL Cancer Res. 39:3571-3574(1979).
RN [12]
RP FUNCTION: BINDS COPPER, NICKEL, AND FATTY ACIDS AS WELL AS, AND
CC BILIRUBIN LESS WELL THAN, SERUM ALBUMIN. ONLY A SMALL PERCENTAGE
CC (LESS THAN 2%) OF THE HUMAN AFP SHOWS ESTROGEN-BINDING PROPERTIES.
CC -!- SUBUNIT: DIMERIC AND TRIMERIC FORMS HAVE BEEN FOUND IN ADDITION
CC TO THE MONOMERIC FORM.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- TISSUE SPECIFICITY: PLASMA. SYNTHESIZED BY THE FETAL LIVER AND

```

CC YOLK SAC.
 CC -|- DEVELOPMENTAL STAGE: OCCURS IN THE PLASMA OF FETUSES MORE THAN 4
 CC WEEKS OLD, REACHES THE HIGHEST LEVELS DURING THE 12TH-16TH WEEK OF
 CC GESTATION, AND DROPS TO TRACE AMOUNTS AFTER BIRTH. THE SERUM LEVEL
 CC IN ADULTS IS USUALLY LESS THAN 40 NG/ML. APP OCCURS ALSO AT HIGH
 CC LEVELS IN THE PLASMA AND ASCITIC FLUID OF ADULTS WITH HEPATOMA.
 CC -|- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
 CC -|- PTM: INDEPENDENT STUDIES SUGGEST HETEROGENEITY OF THE AMINO-
 CC TERMINAL SEQUENCE OF THE MATURE PROTEIN AND OF THE CLEAVAGE SITE
 CC OF THE SIGNAL SEQUENCE.
 CC -|- SIMILARITY: BELONGS TO THE ALB/APP/VDB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M10949; AAA51674.1; -
 CC EMBL; M10950; AAA51675.1; -
 CC EMBL; V01514; CAA24758.1; -
 CC EMBL; M16110; AAB58754.1; -
 CC EMBL; J00077; AAC95396.1; -
 CC EMBL; Z19532; CAA79592.1; -
 CC PIR; A03234; FPHU.
 CC PIR; A26624; A26624.
 CC HSP; P02768; LUOR.
 CC MIM; 104150; -
 CC PRINTS: PR00802; SERUMALBUMIN.
 CC PRINTS: PR00803; FETOPROTEIN.
 CC PROSITE; PS00212; ALBUMIN; 2.
 CC PFAM; PF00273; transport_prot; 1.
 CC Glycoprotein; Albumin; Plasma; Embryo; Repeat; Metal-binding;
 CC Copper; Nickel; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 609 ALPHA-FETOPROTEIN.
 FT METAL 22 22 COPPER AND NICKEL.

FT METAL 22 22
 FT DISULFID 99 114
 FT DISULFID 113 124
 FT DISULFID 148 193
 FT DISULFID 192 201
 FT DISULFID 224 270
 FT DISULFID 269 277
 FT DISULFID 289 303
 FT DISULFID 302 313
 FT DISULFID 384 393
 FT DISULFID 416 462
 FT DISULFID 461 472
 FT DISULFID 485 501
 FT DISULFID 500 511
 FT DISULFID 538 583
 FT DISULFID 582 591
 FT CARBOHYD 251 251
 SQ SEQUENCE 609 AA; 68677 MW; 4D4E45820E1C2D4F CRC64;

Query Match 71.2%; Score 42; DB 1; Length 609;
 Best Local Similarity 50.0%; Pred. No. 1.00e+01;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 80 DQSSGCL 87
 QY 78 EGTSGCL 85
 RESULT 14
 ID LCAT_MOUSE STANDARD; PRT; 438 AA.
 AC P16301;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUL-1999 (Rel. 36, Last annotation update)
 DE PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE PRECURSOR (EC 2.3.1.43)

DE (LECITHIN-CHOLESTEROL ACYLTRANSFERASE) (PHOSPHOLIPID-CHOLESTEROL
 DE ACYLTRANSFERASE).
 GN LCAT.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 90094326.
 RA Warden C.H.; Langner C.A.; Gordon J.I.; Taylor B.A.; McLean J.W.;
 RA Lusis A.J.;
 RT Tissue-specific expression, developmental regulation, and
 RT chromosomal mapping of the lecithin: cholesterol acyltransferase
 RT gene. Evidence for expression in brain and testes as well as liver.;
 RL J. Biol. Chem. 264:21573-21581(1989).
 RN [2]
 RP SEQUENCE OF 1-14 FROM N.A.
 RA Meroni G., Magaretti N., Magnaghi P., Taramelli R.;
 RA Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
 CC -|- FUNCTION: CENTRAL ENZYME IN THE EXTRACELLULAR METABOLISM OF PLASMA
 CC LIPOPROTEINS. AMONG OTHER SUBSTRATES IT ESTERIFIES THE FREE
 CC CHOLESTEROL TRANSPORTED IN PLASMA LIPOPROTEINS.
 CC -|- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + STEROL = STEROL ESTER +
 CC 1-ACYLGLYCEROPHOSPHOCHOLINE (PALMITOYL, OLEOYL, AND LINOLEOYL CAN
 CC BE TRANSFERRED; A NUMBER OF STEROLS, INCLUDING CHOLESTEROL, CAN
 CC ACT AS ACCEPTOR).
 CC -|- ENZYME REGULATION: APOLIPOPROTEIN A-I IS A POTENT ACTIVATOR FOR
 CC THIS ENZYME.
 CC -|- SIMILARITY: PARTIAL WITH LIPASES (PANCREATIC, GASTRIC, HEPATIC,
 CC LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; J05154; AAA39419.1; -
 CC EMBL; X54095; CAA38029.1; -
 CC PIR; A34158; XMSN.
 CC MGD; MGI:96755; LCAT.
 DR PROSITE; PS00120; LIPASE_SER; 1.
 KW Transferase; Acyltransferase; Lipid metabolism; Glycoprotein; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 438 PHOSPHATIDYLCHOLINE-STEROL
 FT ACT_SITE 205 205 ACYLTRANSFERASE.
 FT DISULFID 74 98 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 337 380 BY SIMILARITY.
 FT CARBOHYD 44 44 BY SIMILARITY.
 FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 438 AA; 49765 MW; 2FDD571943433136 CRC64;

Query Match 69.5%; Score 41; DB 1; Length 438;
 Best Local Similarity 71.4%; Pred. No. 1.70e+01;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 203 GHSGLCL 209
 QY 79 GHSGLCL 85
 RESULT 15
 ID LCAT_RABIT STANDARD; PRT; 440 AA.
 AC P53761;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE PRECURSOR (EC 2.3.1.43)
DE (LECITHIN-CHOLESTEROL ACYLTRANSFERASE) (PHOSPHOLIPID-CHOLESTEROL
GN ACYLTRANSFERASE).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND;
RX MEDLINE: 96425081.
RA Murata Y., Maeda E., Yoshino G., Kasuga M.;
RT "Cloning of rabbit LCAT cDNA: Increase in LCAT mRNA abundance in the
RT liver of cholesterol-fed rabbits.";
RL J. Lipid Res. 37:1616-1622(1996).
CC -!- FUNCTION: CENTRAL ENZYME IN THE EXTRACELLULAR METABOLISM OF PLASMA
CC LIPOPROTEINS. AMONG OTHER SUBSTRATES IT ESTERIFIES THE FREE
CC CHOLESTEROL TRANSPORTED IN PLASMA LIPOPROTEINS.
CC -!- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + STEROL -> STEROL ESTER +
CC 1-ACYLGLYCEROPHOSPHOCHOLINE (PALMITOYL, OLEOYL, AND LINOLEOYL CAN
CC BE TRANSFERRED; A NUMBER OF STEROLS, INCLUDING CHOLESTEROL, CAN
CC ACT AS ACCEPTOR).
CC -!- ENZYME REGULATION: APOLIPOPROTEIN A-I IS A POTENT ACTIVATOR FOR
CC THIS ENZYME.
CC -!- SIMILARITY: PARTIAL WITH LIPASES (PANCREATIC, GASTRIC, HEPATIC,
CC LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D13668; BAA02839.1; --
DR PROSITE: PS00120; LIPASE_SER; 1.
KW Transferase; Acyltransferase; Lipid metabolism; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 440
FT FT PHOSPHATIDYLCHOLINE-STEROL
FT FT ACYLTRANSFERASE
FT ACT_SITE 205 205
FT DISULFID 74 98
FT DISULFID 337 380
FT CARBOHYD 44 44
FT CARBOHYD 108 108
FT CARBOHYD 296 296
FT CARBOHYD 408 408
FT SEQUENCE 440 AA; 49559 MW; 1938C5B43BD534AD CRC64;
SQ

Query Match 69.5%; Score 41; DB 1; Length 440;
Best Local Similarity 71.4%; Pred. No. 1.70e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 203 GHSGLCL 209
||: |||
QY 79 GHTSGL 85

Search completed: Wed May 10 12:37:30 2000
Job time : 97 secs.

WPIRLH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 12:37:48 2000; MasPar time 226.86 Seconds
Tabular output not generated. 2.445 Million cell updates/sec

Title: >US-09-376-430-2
Description: (78-85) from US09376430A.pep (12 of 25)
Perfect Score: 59
Sequence: 1 EIGHTSGCL 8

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 19.835; Variance 21.870; scale 0.907

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	51	86.4	204	2	O53254	HYPOTHETICAL 22.6 KD P	1.62e+01
2	50	84.7	237	5	O17636	C33A12.7 PROTEIN (FRAG	2.89e-01
3	50	84.7	237	11	O35385	PROTEIN PHOSPHATASE WI	2.89e-01
4	47	79.7	236	1	O28526	CONSERVED HYPOTHETICAL	1.59e+00
5	45	76.3	251	13	O9W621	NEUT HES1	4.79e+00
6	45	76.3	948	6	O77659	CYCLIC NUCLEOTIDE-GATE	4.79e+00
7	44	74.6	205	1	O27335	CONSERVED PROTEIN.	8.21e+00
8	44	74.6	256	10	O96332	GLYOXALASE II.	8.21e+00
9	44	74.6	291	11	O35502	LECITHIN:CHOLESTEROL A	8.21e+00
10	44	74.6	391	2	O92360	PUTATIVE LIPASE.	8.21e+00
11	44	74.6	563	5	P91015	CODED FOR BY C. ELEGAN	8.21e+00
12	44	74.6	564	5	P91015	CODED FOR BY C. ELEGAN	8.21e+00
13	44	74.6	721	10	O9XFX5	PHENYLALANINE-AMMONIA	8.21e+00
14	44	74.6	902	10	P93024	TRANSCRIPTION FACTOR (8.21e+00
15	44	74.6	902	10	O64965	TRANSCRIPTION FACTOR.	8.21e+00
16	44	74.6	909	5	O76618	Y23H5A.7 PROTEIN.	8.21e+00
17	44	74.6	1239	5	O94902	KU2BANAN.	8.21e+00
18	44	74.6	1275	14	O41977	TEGUMENT PROTEIN/FGARA	8.21e+00
19	44	74.6	1321	11	O70127	BILE SALT EXPORT PUMP.	8.21e+00
20	44	74.6	1321	11	O88331	P-GLYCOPROTEIN SISTER.	8.21e+00

21	43	72.9	142	14	Q81307	POLYPROTEIN PRECURSOR	1.40e+01
22	43	72.9	227	4	O95571	MRNA EXPRESSED IN THYR	1.40e+01
23	43	72.9	253	5	O44462	C09G12.1 PROTEIN.	1.40e+01
24	43	72.9	295	11	O35574	LECITHIN:CHOLESTEROL A	1.40e+01
25	43	72.9	339	2	O9XD85	SECRETED PROTEIN PRECU	1.40e+01
26	43	72.9	741	5	O25257	C-TERMINAL SEQUENCE OF	1.40e+01
27	43	72.9	819	11	O921D8	ZINC FINGER PROTEIN 95	1.40e+01
28	43	72.9	901	11	O63049	RAT ANDROGEN RECEPTOR.	1.40e+01
29	43	72.9	3010	14	P88803	POLYPROTEIN.	1.40e+01
30	43	72.9	3010	14	P90191	CONSERVED HYPOTHETICAL	1.40e+01
31	42	71.2	205	2	O25493	PUTATIVE.	2.36e+01
32	42	71.2	205	2	O92L31	GLUTAMINE SYNTHETASE (2.36e+01
33	42	71.2	296	2	O923J0	GLUTAMINE SYNTHETASE (2.36e+01
34	42	71.2	296	2	O923J3	GLUTAMINE SYNTHETASE (2.36e+01
35	42	71.2	296	2	O923J8	GLUTAMINE SYNTHETASE (2.36e+01
36	42	71.2	297	2	O923J2	GLUTAMINE SYNTHETASE (2.36e+01
37	42	71.2	297	2	O923J1	GLUTAMINE SYNTHETASE (2.36e+01
38	42	71.2	350	14	O9WHF1	HYPOTHETICAL 37.8 KD P	2.36e+01
39	42	71.2	388	2	O92JB9	PUTATIVE OUTER MEMBRAN	2.36e+01
40	42	71.2	388	2	O26031	HYPOTHETICAL 43.0 KD P	2.36e+01
41	42	71.2	396	13	O9YGL3	TAILLESS.	2.36e+01
42	42	71.2	398	2	O51727	PHLC.	2.36e+01
43	42	71.2	1309	10	O64428	REVERSE TRANSCRIPTASE.	2.36e+01
44	41	69.5	906	4	O14526	KIAA0290 (FRAGMENT).	3.95e+01
45	41	69.5	2344	14	Q86114	POLYPROTEIN.	3.95e+01

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	204 AA.
AC	O53254			
DT	O1-JUN-1998	(TREMblrel. 06, Created)		
DT	O1-NOV-1999	(TREMblrel. 06, Last sequence update)		
DE	HYPOTHETICAL 22.6 KD PROTEIN.			
GN	MTV012.21C.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-H37RV;			
RA	OLIVER K., HARRIS D.;			
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-H37RV;			
RA	BARRELL B.G., RAJANDREAM M.A., PARKHILL J., COLE S.T.;			
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-H37RV;			
RX	MEDLINE; 96181548			
RA	PHILIPP W.J., POULET S., EIGMEIER K., PASCOPELLA L.,			
RA	BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,			
RA	COLE S.T.;			
RT	"An integrated map of the genome of the tubercle bacillus,			
RT	Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium			
RT	leprae."			
RL	Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).			
DR	EMBL; AL021287; CAA16092.1;			
DR	PFAM; PF01613; Flavin_Reduct; 1.			
SW	Hypothetical protein.			
KW	SEQUENCE 204 AA; 22587 MW; 69033964 CRC32;			

Query Match 86.4%; Score 51; DB 2; Length 204;
Best Local Similarity 62.5%; Pred. No. 1.62e-01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 37 DGHPSACL 44
QY 78 EIGHTSGCL 85

US-09-376-430-2-12.rspt

Thu May 11 06:49:52 2000

SQ SEQUENCE 757 AA; 86644 MW; DEC9E5C2 CRC32;
 Query Match 84.7%; Score 50; DB 11; Length 757;
 Best Local Similarity 75.0%; Pred. No. 2.89e-01;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 734 EGHASACL 741
 |||||:
 QY 78 EGTSGCL 85

RESULT 4 PRELIMINARY; PRT; 236 AA.
 ID O28526
 AC O28526;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN AFI1748
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE; 98049343.
 RA KLECK H.P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
 RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
 RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRIDES N.C.,
 RA FEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
 RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
 RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLOECK A., ZHOU L.,
 RA OVERBEER R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
 RA COTTON M.D., SPRIGGS T., ARTACH P., KATSEV S.M.,
 RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
 RA VENTER J.C.;
 RI "The complete genome sequence of the hyperthermophilic, sulphate-
 reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997).
 DR EMBL; AE000982; AAB89502.1;
 DR TIGR; AFI1748;
 DR PFAM; PF00753; lactamase_B; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 236 AA; 26635 MW; 9C65C243 CRC32;

Query Match 79.7%; Score 47; DB 1; Length 236;
 Best Local Similarity 85.7%; Pred. No. 1.59e+00;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 140 GHTSGCI 146
 |||||:
 QY 79 GHTSGCL 85

RESULT 5 PRELIMINARY; PRT; 251 AA.
 ID Q9W621
 AC Q9W621;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE NEWT HES1.
 GN NEWT HES1.
 OS Cynops pyrrhogaster (Japanese common newt).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Caudata; Salamandridae; Salamandridae; Cynops.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIMB BLASTOMA; TAZAWA I., UCHIYAMA K., YOSHIZATO K.;
 RA SHIMIZU-NISHIKAWA K.;
 RT "Expression of the helix-loop-helix type negative regulators of
 RT differentiation during limb regeneration in urodele and anuran.";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000893; AAB85756.1; -.

FT	NON_TER	291	..291
SQ	SEQUENCE	291 AA;	335

RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000893; AAB85756.1; -.

```

Query Match          74.6%; Score 44; DB 11; Length 291;
Best Local Similarity 71.4%; Pred. No. 8.21e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 123 GHSVGL 129
QY 79 GHSVGL 85

RESULT 10
ID Q92360 PRELIMINARY; PRT; 391 AA.
AC Q92360;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PUTATIVE LIPASE.
GN SCE8.12C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA SAUNDERS D., HARRIS D.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA BENTLEY S.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE; 97000351.
RA REDENBACH M., KIESER H.M., DENAPATE D., EICHNER A., CULLUM J.,
RA KINASHI H., HOPWOOD D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.COELICOLOR; STRAIN=M145;
RA CHONG P.P.;
RT "Molecular genetic studies on the biosynthesis and regulation of
RT calcium-dependent antibiotic (CDA) production by Streptomyces
RT coelicolor A3(2).";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031654; CAB38590.1; -
DR EMBL; AF124138; AADI8047.1; -
SQ SEQUENCE 391 AA; 41487 MW; EDF9FA86 CRC32;

Query Match          74.6%; Score 44; DB 2; Length 391;
Best Local Similarity 50.0%; Pred. No. 8.21e+00;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 194 DGRVTGCL 201
QY 78 EIGHTSGCL 85

RESULT 11
ID P91015 PRELIMINARY; PRT; 563 AA.
AC P91015;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE CODED FOR BY C. ELEGANS CDNA YK167F5.5.
GN C0168.5A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FULTON L.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FULTON L.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.

Query Match          74.6%; Score 44; DB 5; Length 563;
Best Local Similarity 62.5%; Pred. No. 8.21e+00;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 141 ETHVAGCL 148
QY 78 EIGHTSGCL 85

RESULT 12
ID P91016 PRELIMINARY; PRT; 564 AA.
AC P91016;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE CODED FOR BY C. ELEGANS CDNA YK86F6.5.
GN C0168.5B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FULTON L.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.

```


RC STRAIN-BRISTOL N2;
 RA DU Z., GATTUNG S.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U80439; AAB37642.1; -;
 DR PROSITE; PS00660; BAND_41_1; 1.
 DR PROSITE; PS00661; BAND_41_2; 1.
 DR PFAM; PF00373; Band_41; 1.
 DR PFAM; PF00769; ERM; 1.
 DR PRINTS; PR00661; ERMFAMILY.
 DR PRINTS; PR00935; BAND41
 SQ SEQUENCE 564 AA; 65999 MW; 77C05E44 CRC32;

Query Match 74.6%; Score 44; DB 5; Length 564;
 Best Local Similarity 62.5%; Pred. No. 8.21e+00;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 142 ETHVAGCL 149
 : : : :
 QY 78 EGTSGCL 85

RESULT 13 PRELIMINARY; PRT; 721 AA.
 ID Q9XFX5;
 AC Q9XFX5;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE PHENYLALANINE-AMMONIA LYASE (EC 4.3.1.5).
 GN PALI.
 OS Citrus clementina x Citrus reticulata.
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Sapindales; Rutaceae; Citrus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. FORTUNE; TISSUE-FLAVEDO;
 RA SANCHEZ-BALLESTA M.T., LAFUENTE M.T., ZACARIAS L., GRANELL A.;
 RT "Phenylalanine ammonia-lyase gene expression in response to low
 temperature in mandarin fruit."
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ238753; CAB42793.1; -;
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KW Lyase.
 SQ SEQUENCE 721 AA; 78429 MW; C8E17521 CRC32;

Query Match 74.6%; Score 44; DB 10; Length 721;
 Best Local Similarity 62.5%; Pred. No. 8.21e+00;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 9 NGHQNGCL 16
 : : : :
 QY 78 EGTSGCL 85

RESULT 14 PRELIMINARY; PRT; 902 AA.
 ID P93024;
 AC P93024;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE TRANSCRIPTION FACTOR (IAA24).
 GN MP OR IAA24.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RX MEDLINE; 98151364.
 RA HARDTKE C.S., BERLETH T.;
 RT "The Arabidopsis gene MONOPTEROS encodes a transcription factor
 mediating embryo axis formation and vascular development.";
 RL ENBO J. 17:1405-1411(1998).
 RN [2]
 RP SEQUENCE OF 13-902 FROM N.A.
 RX MEDLINE; 98004476.
 RA HARTER K.J.W., KIM J., THEOLOGIS A.;
 RT "Protein-protein interactions among the Aux/IAA proteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:11786-11791(1997).
 DR EMBL; AF037228; AAC39410.1; -;
 DR EMBL; U79557; AAB92476.1; -;
 DR MENDEL; 24818; Arath; 3165; 24818.
 SQ SEQUENCE 902 AA; 99650 MW; 049F73E1 CRC32;

Query Match 74.6%; Score 44; DB 10; Length 902;
 Best Local Similarity 62.5%; Pred. No. 8.21e+00;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 715 QNHPSGCL 722
 : : : :
 QY 78 EGTSGCL 85

RESULT 15 PRELIMINARY; PRT; 902 AA.
 ID O64965;
 AC O64965;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE TRANSCRIPTION FACTOR.
 GN MP.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RX MEDLINE; 98151364.
 RA HARDTKE C.S., BERLETH T.;
 RT "The Arabidopsis gene MONOPTEROS encodes a transcription factor
 mediating embryo axis formation and vascular development.";
 RL ENBO J. 17:1405-1411(1998).
 DR EMBL; AF037229; AAC60794.1; -;
 DR MENDEL; 28796; Arath; 3165; 28796.
 SQ SEQUENCE 902 AA; 99662 MW; 667EBA47 CRC32;

Query Match 74.6%; Score 44; DB 10; Length 902;
 Best Local Similarity 62.5%; Pred. No. 8.21e+00;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 715 QNHPSGCL 722
 : : : :
 QY 78 EGTSGCL 85

Search completed: Wed May 10 12:41:45 2000
 Job time : 237 secs.

W P S R E H (TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 10 12:50:43 2000; MasPar time 2.80 Seconds
Tabular output not generated. 67.672 Million cell updates/sec

Title: >US-09-376-430-2
Description: (88-95) from US09376430A.pep (13 of 25)
Perfect Score: 53
Sequence: 1 AQRDDIL 8

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:geneseqp

Statistics: Mean 15.330; Variance 40.737; scale 0.376

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	43	81.1	541	W12695	G-protein parathyroid	1.75e+02
2	42	79.2	533	W83317	Mouse Lrp5 protein fro	2.30e+02
3	42	79.2	1240	W90262	A. tygrinum AE2 protei	2.30e+02
4	42	79.2	1279	W98761	H. pylori GHPO 1090 pr	2.30e+02
5	42	79.2	1614	W83312	Mouse Lrp5 protein.	2.30e+02
6	42	79.2	4472	R97245	Virulence gene cluster	2.30e+02
7	41	77.4	130	Y03185	Topoisomerase II bindi	3.01e+02
8	41	77.4	268	R41691	ATGF #2.	3.01e+02
9	41	77.4	434	R13493	P.denitrificans COB B.	3.01e+02
10	41	77.4	479	Y03184	Topoisomerase II bindi	3.01e+02
11	41	77.4	870	W17080	Clarkia breveri S-lina	3.01e+02
12	41	77.4	870	W173485	Linalool synthase prot	3.01e+02
13	40	75.5	154	W28087	Staphylococcus aureus	3.93e+02
14	39	73.6	104	R38894	Sequence of the M1 (ra	5.13e+02
15	39	73.6	155	W27990	Amino acid sequence of	5.13e+02
16	39	73.6	359	W37959	Amino acid sequence of	5.13e+02
17	39	73.6	376	R90541	pJG4-5-CDK-Bp clone #1	5.13e+02
18	39	73.6	380	W09406	Transforming growth fa	5.13e+02
19	39	73.6	422	W31522	Phospholipid scramblas	5.13e+02
20	39	73.6	452	W52121	Trypanosoma lymphocyte	5.13e+02
21	39	73.6	467	W52120	Trypanosoma lymphocyte	5.13e+02
22	39	73.6	493	W52119	Trypanosoma lymphocyte	5.13e+02
23	39	73.6	583	W52118	Trypanosoma lymphocyte	5.13e+02

24	39	73.6	587	1	R88656	N.meningitidis Bz83 Tb	5.13e+02
25	39	73.6	679	1	Y01524	TbpB sequence from Nei	5.13e+02
26	39	73.6	685	1	Y01525	TbpB sequence from Nei	5.13e+02
27	39	73.6	687	1	R32074	Bovine transglutaminas	5.13e+02
28	39	73.6	689	1	Y01523	TbpB sequence from Nei	5.13e+02
29	39	73.6	689	1	R88648	Neisseria meningitidis	5.13e+02
30	39	73.6	999	1	W18511	Human oxygen regulated	5.13e+02
31	39	73.6	999	1	W18512	Rat oxygen regulated p	5.13e+02
32	39	73.6	3685	1	P90290	Human Duchenne muscula	5.13e+02
33	39	73.6	3685	1	P90373	Sequence encoded by hu	5.13e+02
34	38	71.7	38	1	W96856	Nucleic acid binding	6.67e+02
35	38	71.7	201	1	W13274	Human von Willebrand f	6.67e+02
36	38	71.7	227	1	R41277	vWF fragment Met.Tyr50	6.67e+02
37	38	71.7	720	1	W98136	Human transglutaminase	6.67e+02
38	38	71.7	1190	1	R13308	CFTR 3659 del C.	6.67e+02
39	38	71.7	1479	1	R11602	Mutant cystic fibrosis	6.67e+02
40	38	71.7	1480	1	R72852	CFTR.	6.67e+02
41	38	71.7	1480	1	R13894	Cystic fibrosis trans	6.67e+02
42	38	71.7	1480	1	W68012	Human cystic fibrosis	6.67e+02
43	38	71.7	1480	1	W97573	Cystic fibrosis trans	6.67e+02
44	38	71.7	1480	1	R13301	CFTR P574H.	6.67e+02
45	38	71.7	4536	1	W96826	Amino acid sequence of	6.67e+02

ALIGNMENTS

RESULT 1
ID W12695 standard; Protein; 541 AA.
AC W12695;
DT 31-MAY-1997 (first entry)
DE G-protein parathyroid hormone receptor HLTGDG74.
KW G-protein parathyroid hormone receptor; HLTGDG74; parathormone; PTH;
KW calcium; signal transduction; agonist; antagonist; hypocalcaemia;
KW hyperphosphataemia; hypoparathyroidism; chronic tetany;
KW osteoporosis; hypercalcaemia; hypoparathyroidism; hypophosphataemia;
KW kidney stone; nephroliasis; therapy; diagnosis.
OS Homo sapiens.
PN WO9639433-AL.
PD 12-DEC-1996.
PF 05-JUN-1995; U07085.
PR 05-JUN-1995; WO-007085.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Li Y, Rosen CA, Ruben SM, Soppet DR;
DR WPI; 97-043068/04.
DR N-PSDB; T59619.
PT Human G-protein parathyroid hormone receptor, HLTGDG74 - used to
PT identify (ant)agonists, used in the treatment of hypo- or
PT hyper-calcaemia, hypo- or hyper-phosphatemia, kidney stones, etc
PS Claim 9; Fig 1A-E; 62pp; English.
CC A novel 7-transmembrane receptor (W12695) has been identified as a
CC human G-protein parathyroid hormone (PTH) receptor, designated as
CC HLTGDG74.. It shows 48.2% homology to the human PTH receptor. Its
CC amino acid sequence was deduced from a cDNA clone (r19619) isolated
CC from a human T cell lymphoma tissue cDNA library. Recombinant
CC HLTGDG74 can be produced in transformed host cells and used to
CC screen for (ant)agonist cpds. Agonists may be used to prevent or
CC treat e.g. hypocalcaemia, hyperphosphataemia, hypoparathyroidism
CC and chronic tetany by stimulating an increase in serum calcium
CC levels. Antagonists can be used to inhibit the receptor e.g. for
CC the treatment of osteoporosis, hypercalcaemia, hypoparathyroidism,
CC hypophosphataemia, kidney stone, nephroliasis.
SQ Sequence 541 AA;

Query Match 81.1%; Score 43; DB 1; Length 541;
Best Local Similarity 100.0%; Pred. No. 1.75e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 520 QRDDIL 525
QY 90 QRDDIL 95

RESULT 2

ID	W98761	standard; Protein; 1279 AA.
AC	W98761;	
AD	31-MAR-1999	(first entry)
DT	H. Pylori GHPO 1090 protein.	
DE	GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;	
KW	peptic ulcer disease.	
OS	Helicobacter pylori.	
PN	W09843478-A1.	
PD	08-OCT-1998.	
PF	01-APR-1998:	U06371.
PE	29-JUL-1997:	US-902615.
PP	01-APR-1997:	US-833457.
PS	01-JUN-1997:	US-833457.

Aalkjaer C, Benavise MO, Biemesderfer D, Boron WF,
Choi I, Davis BA, Grigorchenko II, Romero MF, Schmitt BM,
Sussman CR:
WPI: 99-059743/05.
New nucleic acid molecules encoding proteins of the Sodium
Bicarbonate Cotransporter (NBC) family - useful for identifying
agents that agonize or antagonise NBC activity and treating
disorders mediated by NBC

ID	W83312 standard; Protein; p614 AA.
AC	W83312;
DT	10-FEB-1999 (first entry)
DE	Mouse Lrp5 protein.
KW	LRP5; LDL-receptor related protein

KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
 KW glomerulonephritis; inflammation; viral infection; osteoporosis;
 KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein.
 OS Mus sp.
 PN WO9846743-A1.
 PD 22-OCT-1998.
 PF 15-APR-1998. G01102.
 PR 05-JUN-1997; US-048740.
 PR 15-APR-1997; US-043553.
 PA (MERI) MERCK & CO INC.
 PA (WELL) WELLCOME TRUST LTD.
 PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW,
 PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
 PI Phillips MS, Todd JA, Twells RCJ;
 DR WPI; 98-594573/50.
 PT New isolated LDL-receptor related protein - used to develop products
 PT for treating, e.g. elevated triglyceride levels, diabetes,
 PT autoimmune disorders, inflammation or Alzheimer's disease
 PS Claim 6; Fig 18c; 200pp; English.
 CC The present invention describes LRP5 (low density lipoprotein (LDL)
 CC receptor related protein, previously designated LRP-3). The present
 CC sequence represents the mouse Lrp5 protein.
 CC Nucleic acid molecules (NAMS) encoding LRP5 can be used for determining,
 CC if an individual is susceptible to insulin dependent diabetes mellitus
 CC (IDDM). The NAMS or proteins can be used for reducing triglyceride levels
 CC in the serum of an individual. Therapies that affect LRP5 may also be
 CC useful in the treatment of autoimmune diseases such as
 CC glomerulonephritis, diseases and disorders involving disruption of
 CC endocytosis and/or antigen presentation, cytokine clearance and/or
 CC inflammation, viral infection, pathogenic bacterial toxin contamination,
 CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
 CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
 CC from the present invention can also be used for detection, diagnosis and
 CC drug screening.
 CC Sequence 1614 AA;

Query Match 79.2%; Score 42; DB 1; Length 1614;
 Best Local Similarity 62.5%; Pred. No. 2.30e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 267 GEORKEIL 274

QY :||:|

QY 88 AEQRDDIL 95

RESULT 6
 ID R97245 standard; Protein; 4472 AA.

AC R97245;
 DT 07-JAN-1997 (first entry)
 DE Virulence gene cluster polypeptide product.
 KW Mutant; adaptation; virulence factor; identification; screening;
 KW vaccine; drugs; infection; treatment.
 OS Salmonella typhimurium.
 FH Key Location/Qualifiers
 FT Region
 FT /note= "All x's in this sequence correspond to
 FT termination codons in the virulence gene
 FT cluster sequence given in T09224."
 FT

FN WO9617951-A2.

PD 13-JUN-1996.

PF 11-DEC-1995; G02875.

PR 09-DEC-1994; GB-024921.

PR 31-JAN-1995; GB-001861.

PR 05-MAY-1995; GB-009239.

PA (RPMs-) RPMs TECHNOLOGY LTD.

PI Holden DW;

DR WPI; 96-287194/29.

PT Identifying virulence genes in microorganisms - by introducing
 PT mutants with insertion inactivated genes into environment and
 PT retrieval and analysis of mutants

PS Claim 5; Figure 11; 131pp; English.

CC A method for identifying a microorganism having a reduced adaptation
 CC to a particular environment comprising the steps of: (1) providing a

CC plurality of microorganisms each of which is independently mutated by
 CC the insertional inactivation of a gene with a nucleic acid comprising
 CC a unique marker sequence so that each mutant contains a different
 CC marker sequence, or clones of the said microorganism; (2) providing
 CC individually a stored sample of each mutant produced by step (1) and
 CC providing individually stored nucleic acid comprising the unique
 CC marker sequence from each individual mutant; (3) introducing a
 CC plurality of mutants produced by step (1) into the said particular
 CC environment and allowing those microorganisms which are able to do so
 CC to grow in the said environment; (4) retrieving microorganisms from
 CC the said environment or a selected part thereof and isolating the
 CC nucleic acid from the retrieved microorganisms; (5) comparing any
 CC marker sequences in the nucleic acid isolated in step (4) to the
 CC unique marker sequence of each individual mutant stored as in step
 CC (2); and (6) selecting an individual mutant which does not contain any
 CC of the marker sequences as isolated in step (4). The products and
 CC methods can be used for identifying virulence genes in microorganisms.
 CC The mutant microorganisms can be used in vaccines or to screen for
 CC drugs which reduce virulence or compounds useful for preventing,
 CC ameliorating or treating infections in animals or plants.
 CC Sequence 4472 AA;

Query Match 79.2%; Score 42; DB 1; Length 4472;
 Best Local Similarity 57.1%; Pred. No. 2.30e+02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1038 DOREQIL 1044

QY :||:|

QY 89 EQRDDIL 95

RESULT 7

ID Y03185 standard; protein; 130 AA.

AC Y03185; (first entry)

DT 16-JUN-1999

DE Topoisomerase II binding protein fragment.

KW Topoisomerase II binding protein; TopBP; anticancer agent.

OS Homo sapiens.

PN J11075856-A.

PD 23-MAR-1999.

PF 17-SEP-1997; 251544.

PR 17-SEP-1997; JP-251544.

PA (CHUS) CHUGAI PHARM CO LTD.

PA (TSUR/) TSURUO T.

DR WPI; 99-257704/22.

PT New topoisomerase II- binding protein - useful as an anticancer

PT agent

PS Disclosure; Page 23; 28pp; Japanese.

CC This sequence is a fragment of the topoisomerase II binding protein
 CC (TopBP) of the invention. The TopBP protein is useful as an anticancer
 CC agent. TopBP can be used as the target molecule for anticancer agent.
 CC Sequence 130 AA;

Query Match 77.4%; Score 41; DB 1; Length 130;
 Best Local Similarity 37.5%; Pred. No. 3.01e+02;
 Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 54 SQEREDVL 61

QY :||:|

QY 88 AEQRDDIL 95

RESULT 8

ID R41691 standard; Protein; 268 AA.

AC R41691;

DT 21-APR-1994 (first entry)

DE AIGF #2.

KW Cellular growth factor; androgen-induced growth factor; AIGF;

KW cell growth; reagent; detection; anticancer.

OS Homo sapiens.

FH Key Location/Qualifiers

FT misc_difference 28.43

FT /note= "Sequence represented as given in the

```

FT protein
FT J05222096-A.
FT 23. .268
FT /note= "Pref. protein"
PN 31-AUG-1993.
PD 28-JAN-1992; 038717.
PF 28-JAN-1992; JP-038717.
PR (SHIO ) SHIONOGI & CO LTD.
PA WPI; 93-309194/39.
DR N-PSDB; 048025.
DR Cell growth factor useful for detection of anticancer substance -
PT has specified amino acid sequence
PS Claim 2; Page 7-8 ; 10pp; Japanese.
CC The sequences given in R41690-91 represent cellular growth factors, esp.
CC androgen-induced growth factor (AIGF). The pref. protein of the
CC invention comprises amino acids His35-Arg215 of sequence #1 or Gln23-
CC Arg268 of sequence #2. These AIGFs are useful as cell growth reagents
CC and for the detection of anticancer substances.
SQ Sequence 268 AA;

Query Match 77.4%; Score 41; DB 1; Length 268;
Best Local Similarity 71.4%; Pred. No. 3.01e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 75 EQRDSVL 81
QY 89 EQRDDIL 95

RESULT 9
ID R13493 standard; Protein; 434 AA.
AC R13493;
DT 25-OCT-1991 (first entry)
DE P.denitrificans COB B.
KW cob gene; corrinoid; descobaltocorrinoid; cor gene.
OS Pseudomonas denitrificans.
PN W09111518-A.
PD 08-AUG-1991.
PF 30-JAN-1991; F00054.
PR 31-JAN-1990; FR-001137.
PA (RHON ) RHONE-POULENC BIOCH.
PI Blanche F, Meron B, Crouzet J, Debussche L, Levy-Schil S;
PI Thibaut D;
DR WPI; 91-252650/34.
DR N-PSDB; Q13284.
PT New polypeptide(s) involved in cobalamin and cobamide
PT biosynthesis - and DNA encoding them, for amplification of
PT cobalamin, esp. coenzyme B12 prodn.
PS Claim 23; Fig 15; 29pp; French.
CC This sequence corresponds to one of 24 polypeptides obtained from
CC P.denitrificans and implicated in the biosynthesis of cobalamines
CC and/or cobamides. It is encoded by part of the 5.4kb
CC ClaI-HindIII-HindIII-HindIII fragment of plasmid pXL157. The
CC plasmid was isolated from a P.denitrificans genomic DNA bank
CC constructed in vector pXL59. Plasmids able to complement
CC P.putida and Agrobacterium tumefaciens cob mutants were identified
CC and their inserts were sequenced. The COB B protein has cobyrinic
CC acid and/or hydrogencobyrinic a,c-diamide synthase activity.
CC See also Q13285-Q13288.
SQ Sequence 434 AA;

Query Match 77.4%; Score 41; DB 1; Length 434;
Best Local Similarity 71.4%; Pred. No. 3.01e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 134 ADHRDDI 140
QY 88 AEQRDDI 94

RESULT 10
ID Y03184 standard; protein; 479 AA.
AC Y03184;

```

```

DT 16-JUN-1999 (first entry)
DE Topoisomerase II binding protein fragment.
KW Topoisomerase II binding protein; TopBP; anticancer agent.
OS Homo sapiens.
PN J11075856-A.
PD 23-MAR-1999.
PF 17-SEP-1997; 251544.
PR 17-SEP-1997; JP-251544.
PA (CHUS ) CHUGAI PHARM CO LTD.
PA (TSURU ) TSURUO T.
DR WPI; 99-257704/22.
DR New Topoisomerase II- binding protein - useful as an anticancer
PT agent
PS Disclosure; Page 21-23; 28pp; Japanese.
CC This sequence is a fragment of the topoisomerase II binding protein
CC (TopBP) of the invention. The TopBP protein is useful as an anticancer
CC agent. TopBP can be used as the target molecule for anticancer agent.
SQ Sequence 479 AA;

Query Match 77.4%; Score 41; DB 1; Length 479;
Best Local Similarity 37.5%; Pred. No. 3.01e+02;
Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 54 SOEREDVL 61
QY 88 AEQRDDIL 95

RESULT 11
ID W17080 standard; Protein; 870 AA.
AC W17080;
DT 09-AUG-1997 (first entry)
DE Clarkia breweri S-linalool synthase.
KW S-linalool synthase; scent; fragrance; aroma; flavour;
KW transgenic plant.
OS Clarkia breweri.
PN W09715584-A2.
PD 01-MAY-1997.
PF 15-OCT-1996; U16482.
PR 12-OCT-1995; US-005146.
PA (UNMI ) UNIV MICHIGAN.
PI Pichersky E;
DR WPI; 97-258947/23.
DR N-PSDB; T68323.
PT Linalool synthase from Clarkia breweri (Onograceae) - an acyclic
PT monoterpene, useful for enhancing the scent production or flavour
PT of plants.
PS Claim 2; Page 83-90; 105pp; English.
CC Clarkia breweri S-linalool synthase (LIS) (W17080) catalyses the
CC cation-dependent and stereoselective conversion of geranyl
CC pyrophosphate to S-linalool, an acyclic monoterpene that is a major
CC component of the plant's scent. LIS is both developmentally and
CC differentially regulated in the various floral organs. Total
CC LIS activity per flower is highest in petals, from which most
CC linalool emission occurs. The enzyme has been purified from the
CC stigmata of C. breweri and can also be obtd. by recombinant
CC expression using an isolated cDNA clone (T68323). LIS can be
CC expressed in transgenic plants to improve scent prodn. (e.g.
CC petunia, rose, carnation etc.) or flavour (e.g. tomato, grape,
CC tea).
SQ Sequence 870 AA;

Query Match 77.4%; Score 41; DB 1; Length 870;
Best Local Similarity 62.5%; Pred. No. 3.01e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 180 SQMRDDIL 187
QY 88 AEQRDDIL 95

RESULT 12
ID W73485 standard; Protein; 870 AA.

```

AC W73485;
 DT 19-FEB-1999 (first entry)
 DE Linalool synthase protein sequence.
 KW Linalool synthase; scent enhancer.
 OS Clarkia breweri.
 PN US5849526-A.
 PD 15-DEC-1998.
 PF 15-OCT-1996; 732192.
 PR 15-OCT-1996; US-732192.
 PA (UNMI) UNIV MICHIGAN.
 PI Pichersky E.
 DR WPI; 99-069727/06.
 DR N-PSDB; V08878.
 PT DNA encoding Clarkia linalool synthase protein - for producing
 PT recombinant protein or transgenic plants
 PS Claim 2; Column 33-40; 30pp; English.
 CC This sequence is encoded by the nucleic acid of the invention, and is the
 CC linalool synthase protein of Clarkia breweri. The DNA is used to produce
 CC recombinant Clarkia breweri S-linalool synthase or to produce transgenic
 CC plants that express Clarkia breweri S-linalool synthase (especially for
 CC the production of enhanced scent and taste in plants).
 SQ Sequence 870 AA;
 Query Match 77.4%; Score 41; DB 1; Length 870;
 Best Local Similarity 62.5%; Pred. No. 3.01e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 180 SQNRDIL 187
 : : : : :
 QY 88 AEQRDDIL 95
 : : : : :
 RESULT 13
 ID W28087 standard; Protein; 154 AA.
 AC W28087;
 DT 01-SEP-1998 (first entry)
 DE Staphylococcus aureus protein of unknown function.
 KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
 KW Staphylococcal gene; regulatory element; bacterial gene expression;
 KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
 KW toxic shock syndrome.
 OS Staphylococcus aureus.
 PN W09730070-A1.
 PD 21-AUG-1997.
 PF 19-FEB-1997; U02318.
 PR 20-FEB-1996; US-011888.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
 DR WPI; 97-424969/39.
 DR N-PSDB; T84024.
 PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
 PT to isolate antimicrobial compounds, and in vaccines against S.
 PT aureus infection
 PS Claim 6; Page 456-457; 989pp; English.
 CC The present sequence represents a Staphylococcus aureus protein of
 CC unknown function. The DNA sequence was isolated from a library of
 CC clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can
 CC be used in the construction of ribozymes and antisense sequences to
 CC control the expression of Staphylococcal genes. The DNA sequence is
 CC also useful as a source of regulatory elements for the control of
 CC bacterial gene expression. The present protein may be used to produce
 CC vaccines to enable a host to produce specific antibodies with
 CC antibacterial action. These vaccines and antibodies would protect
 CC a host against invasion by S. aureus, and conditions relating to
 CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
 CC skin syndrome, and toxic shock syndrome.
 SQ Sequence 154 AA;
 Query Match 75.5%; Score 40; DB 1; Length 154;
 Best Local Similarity 71.4%; Pred. No. 3.93e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 50 ABEREDI 56
 : : : : :
 QY 88 AEQRDDI 94
 : : : : :
 RESULT 14
 ID R38894 standard; Protein; 104 AA.
 AC R38894;
 DT 01-DEC-1993 (first entry)
 DE Sequence of the M1 (rat) transcription cofactor ("DCoH")
 DE protein.
 KW Transcription cofactor; intermolecular interaction.
 OS Balaenoptera acutorostrata.
 PN W09312133-A.
 PD 24-JUN-1993.
 PF 15-DEC-1992; U10866.
 PR 17-DEC-1991; US-809436.
 PA (STRD) UNIV LELAND STANFORD JUNIOR.
 PI Crabtree GR, Mendel DB;
 DR WPI; 93-214087/26.
 DR N-PSDB; Q43421.
 PT Transcriptional cofactors - in which intermolecular interactions
 PT between transcription factors are enhanced, useful to regulate
 PT co-ordinated gene expression, e.g. for gene therapy
 PS Disclosure; Page 60; 79pp; English.
 CC The transcriptional cofactors (DCoH) are relatively small peptides
 CC having a domain of between 20-50 AAs which is lipophilic. They
 CC normally exist as dimers in solution and are conserved across
 CC mammalian species. The cofactors bind to at least one of the HNF-1-
 CC alpha dimer and HNF-1-beta and their heterodimers.
 SQ Sequence 104 AA;
 Query Match 73.6%; Score 39; DB 1; Length 104;
 Best Local Similarity 62.5%; Pred. No. 5.13e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Db 10 ABERDOLL 17
 : : : : :
 QY 88 AEQRDDIL 95
 : : : : :
 RESULT 15
 ID W27990 standard; Protein; 155 AA.
 AC W27990;
 DT 26-AUG-1998 (first entry)
 DE Amino acid sequence of dehydrogenase/reductase-like protein.
 KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
 KW Staphylococcal gene; regulatory element; bacterial gene expression;
 KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
 KW toxic shock syndrome.
 OS Staphylococcus aureus.
 PN W09730070-A1.
 PD 21-AUG-1997.
 PF 19-FEB-1997; U02318.
 PR 20-FEB-1996; US-011888.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
 DR WPI; 97-424969/39.
 DR N-PSDB; T83949.
 PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
 PT to isolate antimicrobial compounds, and in vaccines against S.
 PT aureus infection
 PS Claim 6; Page 392; 989pp; English.
 CC The present sequence represents a Staphylococcus aureus protein, that,
 CC based on homology is believed to be a dehydrogenase/reductase-like
 CC protein. The DNA sequence was isolated from a
 CC library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA
 CC sequence can be used in the construction of ribozymes and antisense
 CC sequences to control the expression of Staphylococcal genes. The DNA
 CC sequence is also useful as a source of regulatory elements for the
 CC control of bacterial gene expression. The present protein may be used
 CC to produce vaccines to enable a host to produce specific antibodies

CC with antibacterial action. These vaccines and antibodies would protect
 CC a host against invasion by S. aureus, and conditions relating to
 CC staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
 CC skin syndrome, and toxic shock syndrome.
 SQ Sequence 155 AA;

Query Match 73.6%; Score 39; DB 1; Length 155;
 Best Local Similarity 71.4%; Pred. No. 5.13e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 13 QQRDDIF 19
 QY 89 EQRDDIL 95

Search completed: Wed May 10 12:50:50 2000
 Job time : 7 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 12:51:08 2000; MasPar time 45.01 Seconds
Tabular output not generated. 2.303 Million cell updates/sec

Title: >US-09-376-430-2
Description: (88-95) from US09376430A.pep (13 of 25)
Perfect Score: 53
Sequence: 1 AEQRDDIL 8

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 14.448; Variance 39.000; scale 0.370

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	43	81.1	234	2	US-07-690-Sequence 2, Applicatio	8.63e+01
2	43	81.1	344	2	US-08-846-Sequence 90, Applicati	8.63e+01
3	43	81.1	541	3	PCT-US95-0Sequence 2, Applicatio	8.63e+01
4	42	79.2	178	2	US-08-637-Sequence 225, Applicat	1.14e+02
5	41	77.4	870	2	US-08-732-Sequence 2, Applicatio	1.49e+02
6	40	75.5	335	2	US-08-761-Sequence 2, Applicatio	1.96e+02
7	39	73.6	104	1	US-08-156-Sequence 4, Applicatio	2.56e+02
8	39	73.6	104	3	PCT-US92-1Sequence 4, Applicatio	2.56e+02
9	39	73.6	104	1	US-08-340-Sequence 4, Applicatio	2.56e+02
10	39	73.6	132	1	US-08-208-Sequence 2, Applicatio	2.56e+02
11	39	73.6	359	2	US-08-713-Sequence 2, Applicatio	2.56e+02
12	39	73.6	370	2	US-08-846-Sequence 81, Applicati	2.56e+02
13	39	73.6	376	1	US-08-253-Sequence 33, Applicati	2.56e+02
14	39	73.6	409	2	US-08-576-Sequence 48, Applicati	2.56e+02
15	39	73.6	431	2	US-08-576-Sequence 54, Applicati	2.56e+02
16	39	73.6	452	2	US-08-586-Sequence 18, Applicati	2.56e+02
17	39	73.6	467	2	US-08-586-Sequence 17, Applicati	2.56e+02
18	39	73.6	493	2	US-08-686-Sequence 5, Applicatio	2.56e+02
19	39	73.6	493	2	US-08-686-Sequence 16, Applicati	2.56e+02
20	39	73.6	999	2	US-08-770-Sequence 1, Applicatio	2.56e+02
21	39	73.6	999	2	US-08-770-Sequence 3, Applicatio	2.56e+02
22	38	71.7	38	2	US-09-066-Sequence 11, Applicati	3.35e+02
23	38	71.7	38	2	US-08-555-Sequence 11, Applicati	3.35e+02

24	38	71.7	49	2	US-08-284-Sequence 9, Applicatio	3.35e+02
25	38	71.7	83	2	US-08-744-Sequence 2, Applicatio	3.35e+02
26	38	71.7	83	1	US-08-555-Sequence 2, Applicatio	3.35e+02
27	38	71.7	83	2	US-08-558-Sequence 2, Applicatio	3.35e+02
28	38	71.7	93	2	US-08-797-Sequence 10, Applicati	3.35e+02
29	38	71.7	226	2	US-08-347-Sequence 4, Applicatio	3.35e+02
30	38	71.7	294	2	US-08-284-Sequence 8, Applicatio	3.35e+02
31	38	71.7	320	2	US-07-841-Sequence 15, Applicati	3.35e+02
32	38	71.7	351	1	US-08-468-Sequence 16, Applicati	3.35e+02
33	38	71.7	442	4	5470718-3Patent No. 5470718	3.35e+02
34	38	71.7	442	4	5470718-2Patent No. 5470718	3.35e+02
35	38	71.7	442	1	US-08-413-Sequence 22, Applicati	3.35e+02
36	38	71.7	463	2	US-08-670-Sequence 18, Applicati	3.35e+02
37	38	71.7	637	1	US-08-350-Sequence 10, Applicati	3.35e+02
38	38	71.7	637	3	PCT-US94-0Sequence 10, Applicati	3.35e+02
39	38	71.7	836	2	US-08-812-Sequence 2, Applicatio	3.35e+02
40	38	71.7	1480	4	5240846-5Patent No. 5240846	3.35e+02
41	38	71.7	1480	2	US-08-455-Sequence 14, Applicati	3.35e+02
42	38	71.7	1480	1	US-08-136-Sequence 2, Applicatio	3.35e+02
43	38	71.7	1480	1	US-08-135-Sequence 2, Applicatio	3.35e+02
44	38	71.7	1480	3	PCT-US93-1Sequence 2, Applicatio	3.35e+02
45	38	71.7	2713	3	PCT-US96-0Sequence 1, Applicatio	3.35e+02

ALIGNMENTS

RESULT 1
ID US-07-690-192-2 STANDARD; PRT; 234 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 2, Application US/07690192
XX
CC Sequence 2, Application US/07690192
CC Patent No. 5919650
CC GENERAL INFORMATION:
CC APPLICANT: Montano, Ximena
CC APPLICANT: Barbacid, Mariano
CC TITLE OF INVENTION: METHOD FOR INACTIVATION OF PROTEIN
CC TITLE OF INVENTION: FUNCTION
CC NUMBER OF SEQUENCES: 4
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Bristol-Myers Squibb Company
CC STREET: P.O. Box 4000
CC CITY: Princeton
CC STATE: New Jersey
CC COUNTRY: U.S.A.
CC ZIP: 08543-4000
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/690,192
CC FILING DATE: 19910422
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Bogden, James M
CC REGISTRATION NUMBER: 32,962
CC REFERENCE/DOCKET NUMBER: DC13
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (609) 921-4163
CC TELEFAX: (609) 921-4526
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 234 amino acids
CC TYPE: AMINO ACID
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 234 AA; 25723 MW; 318574 CN;

```

Query Match      81.1%; Score 43; DB 2; Length 234;
Best Local Similarity 62.5%; Pred. NO. 8.63e+01;
Matches      5; Conservative      2; Mismatches      1; Indels      0; Gaps      0;

Db      173 SEQRDGLV 180
      :||||:|
QY      88 AEQRDDIL 95

RESULT      2
ID      US-08-846-762-90      STANDARD;      PRT;      344 AA.
XX
AC      xxxxxx
XX
XX
DT
XX
DE
DE
XX
XX
Sequence 90, Application US/08846762A
XX
Sequence 90, Application US/08846762A
CC      Patent NO. 5994072
CC      GENERAL INFORMATION:
CC      APPLICANT: Lam, Joseph S.
CC      APPLICANT: Burrows, Lori
CC      APPLICANT: Charter, Deborah
CC      APPLICANT: de Kievit, Teresa
CC      TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
CC      TITLE OF INVENTION: Of O-Antigen in Pseudomonas Aeruginosa
CC      FILE REFERENCE: 6580-089
CC      CURRENT APPLICATION NUMBER: US/08/846,762A
CC      CURRENT FILING DATE: 1997-04-30
CC      NUMBER OF SEQ ID NOS: 100
CC      SOFTWARE: PatentIn Ver. 2.0
CC      SEQ ID NO 90
CC      LENGTH: 344
CC      TYPE: PRT
CC      ORGANISM: Pseudomonas aeruginosa
CC      SEQUENCE      344 AA; 38131 MW; 580679 CN;

Query Match      81.1%; Score 43; DB 2; Length 344;
Best Local Similarity 100.0%; Pred. NO. 8.63e+01;
Matches      6; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Db      230 QRDDIL 235
      :|||||
QY      90 QRDDIL 95

RESULT      3
ID      PCT-US95-07085-2      STANDARD;      PRT;      541 AA.
XX
AC      xxxxxx
XX
XX
DT
XX
DE
XX
Sequence 2, Application PC/TUS9507085
XX
Sequence 2, Application PC/TUS9507085
CC      GENERAL INFORMATION:
CC      APPLICANT: Soppet, Daniel R
CC      APPLICANT: Yi, Li
CC      APPLICANT: Rosen, Craig A
CC      APPLICANT: Ruben, Steven
CC      TITLE OF INVENTION: G-Protein Parathyroid Hormone Receptor
CC      TITLE OF INVENTION: HLTDG74
CC      NUMBER OF SEQUENCES:      8
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
CC      ADDRESSEE: Stewart & Olstein
CC      STREET: 6 Becker Farm Road
CC      CITY: Roseland
CC      STATE: NJ
CC      COUNTRY: USA

```

```
CC REFERENCE/DOCKET NUMBER: RPMS 101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (404) 873-8794
CC TELEFAX: (404) 873-8795
CC INFORMATION FOR SEQ ID NO: 225:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 178 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC SEQUENCE 178 AA; 21191 MW; 147558 CN;

Query Match 79.2%; Score 42; DB 2; Length 178;
Best Local Similarity 57.1%; Pred. No. 1.14e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 159 DQEQIL 165
QY 89 EQRDDIL 95

RESULT 5
ID US-08-732-192A-2 STANDARD; PRT; 870 AA.
XX
AC xxxxxx
XX
DT
DE
Sequence 2, Application US/08732192A
XX
Sequence 2, Application US/08732192A
Patent No. 5849526
CC GENERAL INFORMATION:
CC APPLICANT: Pichersky, Eran
CC TITLE OF INVENTION: USE OF LINALOOL SYNTHASE IN GENETIC
CC ENGINEERING OF SCENT PRODUCTION
CC NUMBER OF SEQUENCES: 2
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ARNOLD, WHITE & DURKEE
CC STREET: P.O. Box 4433
CC CITY: Houston
CC STATE: Texas
CC COUNTRY: USA
CC ZIP: 77210-4433
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/732,192A
CC FILING DATE: 15-NOV-1996
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 60/005,146
CC FILING DATE: 25-OCT-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: HIGHLANDER, STEVEN L.
CC REGISTRATION NUMBER: 37,642
CC REFERENCE/DOCKET NUMBER: UMIC:015
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (512) 418-3000
CC TELEFAX: (713) 789-2679
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 870 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 870 AA; 99839 MW; 4000805 CN;

Query Match 77.4%; Score 41; DB 2; Length 870;
Best Local Similarity 62.5%; Pred. No. 1.49e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 180 SQNRDIL 187
QY 88 AEQRDIL 95

RESULT 6
ID US-08-761-344-2 STANDARD; PRT; 335 AA.
XX
AC xxxxxx
XX
DT
DE
Sequence 2, Application US/08761344
XX
Sequence 2, Application US/08761344
Patent No. 5912154
CC GENERAL INFORMATION:
CC APPLICANT: Ferro-No. 5912154ick, Susan
CC APPLICANT: Jiang, Yu
CC TITLE OF INVENTION: GERANYLGERANYL DIPHOSPHATE SYNTHASE
CC ENGINEERING OF SCENT PRODUCTION
CC NUMBER OF SEQUENCES: 5
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Sheridan Ross P.C.
CC STREET: 1700 Lincoln Street
CC CITY: Denver
CC STATE: CO
CC COUNTRY: U.S.A.
CC ZIP: 80203
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/761,344
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Connell, Gary J.
CC REGISTRATION NUMBER: 30,020
CC REFERENCE/DOCKET NUMBER: 3161-14
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 303/863-9700
CC TELEFAX: 303/862-0223
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 335 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 335 AA; 38651 MW; 567718 CN;

Query Match 75.5%; Score 40; DB 2; Length 335;
Best Local Similarity 50.0%; Pred. No. 1.96e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 251 TEQHNEIL 258
QY 88 AEQRDIL 95

RESULT 7
ID US-08-156-383-4 STANDARD; PRT; 104 AA.
XX
AC xxxxxx
XX
DT
DE
Sequence 2, Application US/08761344
```

DE Sequence 4, Application US/08156383
XX Sequence 4, Application US/08156383
CC Patent No. 5403712
CC GENERAL INFORMATION:
CC APPLICANT: Crabtree, Gerald R
CC APPLICANT: Mendel, Dirk B
CC TITLE OF INVENTION: Transcriptional cofactor regulation of
CC TITLE OF INVENTION: expression
CC NUMBER OF SEQUENCES: 14
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Bertram I. Rowland
CC STREET: 4 Embarcadero Center, Suite 3400
CC CITY: San Francisco
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94111
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/156,383
CC FILING DATE: 17-DEC-1991
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/809,436
CC FILING DATE: 17-DEC-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Rowland, Bertram I
CC REGISTRATION NUMBER: 20,015
CC REFERENCE/DOCKET NUMBER: A55854/BIR-ST
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 781-1989
CC TELEFAX: (415) 398-3249
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 104 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 104 AA; 12000 MW; 54333 CN;
Query Match 73.6%; Score 39; DB 1; Length 104;
Best Local Similarity 62.5%; Pred. No. 2.56e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 10 AERDQLL 17
QY 88 AEQRDDIL 95
RESULT 8
ID PCT-US92-10866-4 STANDARD; PRT; 104 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 4, Application PC/TUS9210866
CC GENERAL INFORMATION:
CC APPLICANT: The Board of Trustees of the Leland Stanford
CC APPLICANT: Junior University
CC TITLE OF INVENTION: Transcriptional cofactor regulation of
CC TITLE OF INVENTION: expression
CC NUMBER OF SEQUENCES: 14
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Bertram I. Rowland
CC STREET: 4 Embarcadero Center, Suite 3400
CC CITY: San Francisco
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94111
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/156,383
CC FILING DATE: 17-DEC-1991
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/809,436
CC FILING DATE: 17-DEC-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Rowland, Bertram I
CC REGISTRATION NUMBER: 20,015
CC REFERENCE/DOCKET NUMBER: A55854/BIR-ST
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 781-1989
CC TELEFAX: (415) 398-3249
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 104 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 104 AA; 12000 MW; 54333 CN;
Query Match 73.6%; Score 39; DB 1; Length 104;
Best Local Similarity 62.5%; Pred. No. 2.56e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 10 AERDQLL 17
QY 88 AEQRDDIL 95
RESULT 8
ID PCT-US92-10866-4 STANDARD; PRT; 104 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 4, Application US/08340136
CC GENERAL INFORMATION:
CC APPLICANT: Crabtree, Gerald R
CC APPLICANT: Mendel, Dirk B
CC TITLE OF INVENTION: Transcriptional cofactor regulation of
CC TITLE OF INVENTION: expression
CC NUMBER OF SEQUENCES: 14
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Bertram I. Rowland
CC STREET: 4 Embarcadero Center, Suite 3400
CC CITY: San Francisco
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94111
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/340,136
CC FILING DATE: 15-NOV-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/809,436

RESULT 12

ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/253,155A
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MIT-028

ATTORNEY/AGENT INFORMATION:
NAME: Dianne Casuto
REGISTRATION NUMBER: P-40,943
REFERENCE/DOCKET NUMBER: 5857, US.O1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847) 938-3137
TELEFAX: (847) 938-2623
TELEX: ⁺₊
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 409 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

CC MOLECULE TYPE: NO. 5998194e
SQ SEQUENCE 409 AA; 45046 MW; 850955 CN;

Query Match 73.6%; Score 39; DB 2; Length 409;
Best Local Similarity 71.4%; Pred. No. 2.56e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 14 EERADIL 20
|:| |
QY 89 EQRDDIL 95

RESULT 15
ID US-08-576-626A-54 STANDARD; PRT; 431 AA.

XX
AC
XX
XX
DT
XX

xxxxxx

Sequence 54, Application US/08576626A

Sequence 54, Application US/08576626A

Patent No. 5998194

GENERAL INFORMATION:

APPLICANT: Summers, R.G.

APPLICANT: Katz, L.

APPLICANT: Donadio, S.

APPLICANT: Staver, M.J.

TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR

TITLE OF INVENTION: BIOSYNTHESIS GENES

NUMBER OF SEQUENCES: 60

CORRESPONDENCE ADDRESS:

ADDRESSEE: Abbott Laboratories

STREET: 100 Abbott Park Road

CITY: Abbott Park

STATE: Illinois

COUNTRY: USA

ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/576,626A

FILING DATE: 21-DEC-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Dianne Casuto

REGISTRATION NUMBER: P-40,943

REFERENCE/DOCKET NUMBER: 5857.US.01

TELECOMMUNICATION INFORMATION:

TELEPHONE: (847) 938-3137

TELEFAX: (847) 938-2623

TELEX:

INFORMATION FOR SEQ ID NO: 54:

SEQUENCE CHARACTERISTICS:

LENGTH: 431 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 5998194e

SEQUENCE 431 AA; 46555 MW; 893911 CN;

Query Match 73.6%; Score 39; DB 2; Length 431;

Best Local Similarity 75.0%; Pred. No. 2.56e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 174 AERRDDAL 181

|:| |

QY 88 AQRDDIL 95

Search completed: Wed May 10 12:52:03 2000
Job time : 55 secs.

W P S R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 10 12:50:18 2000; MasPar time 3.93 Seconds
Tabular output not generated. 96.043 Million cell updates/sec

Title: >US-09-376-430-2
Description: (88-95) from US09376430A.pep (13 of 25)
Perfect Score: 53
Sequence: 1 AQRDDIL 8

Scoring table: PAM 150
Gap 11

Searched: 142080 segs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 21.941; Variance 24.985; scale 0.878

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	49	92.5	132	2	PC4020	hypothetical 132 prot
2	47	88.7	321	2	A44230	phthalate dioxygenase
3	47	88.7	454	2	D69177	hypothetical protein
4	46	86.8	135	2	E69041	hypothetical protein
5	45	84.9	1612	2	S59969	DNA topoisomerase (AT
6	45	84.9	3225	2	I52300	giantin - human
7	45	84.9	3259	1	A56539	giantin - human
8	44	83.0	271	2	B50176	hypothetical protein
9	44	83.0	288	2	I55184	repA protein - Escher
10	44	83.0	2163	2	S50675	pre-mRNA splicing hel
11	43	81.1	106	1	K1RTA	Ig kappa chain C regi
12	43	81.1	214	1	T44001	nef protein - human i
13	43	81.1	218	1	ASLJBR	nef protein - human i
14	43	81.1	251	1	QCQV2P	coat protein - potato
15	43	81.1	309	2	C72589	hypothetical protein
16	43	81.1	341	2	B72097	diptide transporter
17	43	81.1	344	2	S51264	probable galactosyltr
18	43	81.1	351	2	G71372	probable ribonucleosi
19	43	81.1	355	2	G64034	hypothetical protein
20	43	81.1	450	2	A38171	L-lysine 6-transamina
21	43	81.1	452	2	F69294	DNA repair protein RA
22	43	81.1	464	2	S52838	dhIR protein - Xantho
23	43	81.1	468	2	H69500	heme biosynthesis pro

24 43 81.1 703 2i T15234 hypothetical protein 4.85e+01
25 43 81.1 1030 2 S11034 gene P1 protein - fru 4.85e+01
26 43 81.1 1234 2 A34911 band 3-related protei 4.85e+01
27 43 81.1 1237 2 A31789 band 3-related protei 4.85e+01
28 43 81.1 1252 2 D71810 probable type II DNA 4.85e+01
29 42 79.2 133 2i S69803 hypothetical protein 7.74e+01
30 42 79.2 339 2 A43542 lymphocyte antigen WP 7.74e+01
31 42 79.2 339 2 A43512 lymphocyte-specific p 7.74e+01
32 42 79.2 419 2 S46152 probable membrane pro 7.74e+01
33 42 79.2 561 1 VHXPNP major structural nucl 7.74e+01
34 42 79.2 605 2 E69153 cadmium efflux ATPase 7.74e+01
35 42 79.2 844 2 I50159 anion transporter - c 7.74e+01
36 42 79.2 865 2 A25104 band 3 protein, noner 7.74e+01
37 42 79.2 922 2 A30816 band 3 anion transpor 7.74e+01
38 42 79.2 1237 2 A56764 band 3-related protei 7.74e+01
39 42 79.2 1240 2 S21086 anion exchange protei 7.74e+01
40 42 79.2 1264 2 S64146 probable membrane pro 7.74e+01
41 42 79.2 1279 2 E64709 type IIS restriction 7.74e+01
42 42 79.2 1584 2 T15822 kinesin-like protein 7.74e+01
43 42 79.2 2541 2 S11661 talin - mouse 7.74e+01
44 42 79.2 2569 2 T14164 peptidase synthetase ho 7.74e+01
45 42 79.2 4574 2 G02520 plectin - human 7.74e+01

ALIGNMENTS

RESULT 1

ENTRY PC4020 #type fragment
TITLE hypothetical 132 protein - Vibrio cholerae (fragment)
ORGANISM #formal_name Vibrio cholerae
DATE 21-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 19-Oct-1995
ACCESSIONS PC4020
REFERENCE JC4066
#authors Stroemer, U.H.; Karageorgos, L.E.; Morona, R.; Manning, P.A.
#journal Gene (1995) 155:67-72
#title In Vibrio cholerae serogroup O1, rfaD is closely linked to the rfb operon.
#cross-references MUID:95212931
#accession PC4020

#molecule_type DNA
##residues 1-132 ##label STR
##cross-references EMBL:X59554
SUMMARY #length 132 #checksum 4186

Query Match 92.5%; Score 49; DB 2; Length 132;
Best Local Similarity 62.5%; Pred. No. 2.47e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 44 SEQREIL 51
QY 88 AQRDDIL 95

RESULT 2

ENTRY A44230 #type complete
TITLE phthalate dioxygenase reductase (EC 1.18.1.1) - Pseudomonas cepacia
ORGANISM #formal_name Pseudomonas cepacia
DATE 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 19-Dec-1998
ACCESSIONS A44230
REFERENCE A44230

#authors Correll, C.C.; Battie, C.J.; Ballou, D.P.; Ludwig, M.L.
#journal Science (1992) 258:1604-1610
#title Phthalate dioxygenase reductase: a modular structure for electron transfer from pyridine nucleotides to [2Fe-2S].
#cross-references MUID:93088078
#accession A44230
##status preliminary
#molecule_type protein
##residues 1-321 ##label COR
#note sequence extracted from NCBI backbone (NCBIP:120256)

```

CLASSIFICATION #superfamily phthalate dioxygenase reductase; cytochrome-b5
                reductase homology; ferredoxin [2Fe-2S] homology
KEYWORDS        oxidoreductase
FEATURE         #domain ferredoxin [2Fe-2S] homology #label FER
257-309         #length 321 #molecular-weight 34785 #checksum 9114
SUMMARY

Query Match      88.7%; Score 47; DB 2; Length 321;
Best Local Similarity 62.5%; Pred. No. 6.87e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 289 ADRDDVL 296
QY 88 AEQRDDIL 95
|||||

RESULT 3
ENTRY #type complete
TITLE hypothetical protein MTH584 - Methanobacterium
ORGANISM thermoautotrophicum (strain Delta H)
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1999
ACCESSIONS D69177
REFERENCE A69000
#authors Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
Dubois, J.; Algrede, T.; Bashirzadeh, R.; Blakely, D.;
Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicair, R.;
Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; Caruso,
A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.;
Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
J.; Reeve, J.N.
#journal J. Bacteriol. (1997) 179:7135-7155
#title Complete genome sequence of Methanobacterium
thermoautotrophicum Delta H: functional analysis and
comparative genomics.
#cross-references MUID:98037514
#accession D69177
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-454 #label MTH
#cross-references GB:AE000840; GB:AE000666; NID:g2621653;
PIDN:AAB85090.1; PID:g2621662
#experimental_source strain Delta H
GENETICS
#gene MTH584
CLASSIFICATION #superfamily conserved hypothetical protein M1111
SUMMARY #length 454 #molecular-weight 51345 #checksum 803

Query Match      88.7%; Score 47; DB 2; Length 454;
Best Local Similarity 85.7%; Pred. No. 6.87e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 152 EQRDDVL 158
QY 89 EQRDDIL 95
|||||

RESULT 4
ENTRY #type complete
TITLE hypothetical protein MTH1313 - Methanobacterium
thermoautotrophicum (strain Delta H)
ORGANISM thermoautotrophicum thermoautotrophicum
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
05-Jun-1998
ACCESSIONS E69041
REFERENCE A69000
#authors Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;

```

```

Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicair, R.;
Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; Caruso,
A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.;
Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
J.; Reeve, J.N.
#journal J. Bacteriol. (1997) 179:7135-7155
#title Complete genome sequence of Methanobacterium
thermoautotrophicum Delta H: functional analysis and
comparative genomics.
#cross-references MUID:98037514
#accession E69041
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-135 #label MTH
#cross-references GB:AE000895; GB:AE000666; NID:g2622403; PID:g2622418
#experimental_source strain Delta H
GENETICS
#gene MTH1313
SUMMARY #length 135 #molecular-weight 15015 #checksum 8321

Query Match      86.8%; Score 46; DB 2; Length 135;
Best Local Similarity 85.7%; Pred. No. 1.13e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 45 EHRDDIL 51
QY 89 EQRDDIL 95
|||||

RESULT 5
ENTRY #type complete
TITLE DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) isoform
beta - Chinese hamster
ALTERNATE_NAMES DNA topoisomerase II isoform beta; DNA-gyrase
ORGANISM #formal_name Cricetus griseus #common_name Chinese hamster
DATE 15-Feb-1996 #sequence_revision 13-Mar-1997 #text_change
16-Jul-1999
ACCESSIONS S59969; S54154
REFERENCE S59969
#authors Deredre, S.; Frey, S.; Delaporte, C.; Jacquemin-Sablon, A.
#journal Biochim. Biophys. Acta (1995) 1264:178-182
#title Cloning and characterization of full-length cDNAs coding for
the DNA topoisomerase II beta from Chinese hamster lung
cells sensitive and resistant to 9-OH-ellipticine.
#cross-references MUID:96085121
#accession S59969
#status translation not shown
#molecule_type mRNA
#residues 1-1612 #label DBR
#cross-references EMBL:X86455; NID:g790987; PIDN:CAA60173.1;
PID:g790988
#experimental_source lung
CLASSIFICATION #superfamily eukaryotic type II DNA topoisomerase; phage T4
DNA topoisomerase (ATP-hydrolyzing) medium chain homology
KEYWORDS ATP; DNA binding; DNA replication; isomerase; nucleus
FEATURE 697-927
SUMMARY #domain phage T4 DNA topoisomerase (ATP-hydrolyzing)
medium chain homology #label T4T
#length 1612 #molecular-weight 182074 #checksum 5332

Query Match      84.9%; Score 45; DB 2; Length 1612;
Best Local Similarity 62.5%; Pred. No. 1.85e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1189 AQERDIL 1196
QY 88 AEQRDDIL 95
|||||

RESULT 6
ENTRY #type complete

```

```

TITLE          giantin - human
ALTERNATE_NAMES gcp372
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change
                20-Aug-1999
ACCESSIONS      152300
REFERENCE        152300
#authors        Sohda, M.; Misumi, Y.; Fujiwara, T.; Nishioaka, M.; Ikehara,
#journal         Y.
#journal         Biochem. Biophys. Res. Commun. (1994) 205:1399-1408
#title          Molecular cloning and sequence analysis of a human 372-kDa
#title          protein localized in the Golgi complex.
#cross-references MUID:95100974
#accession      152300
##status        preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues      1-3225 ##label RES
##cross-references GB:D25342; NID:g662389; PIDN:BAA05025.1; PID:g808869
CLASSIFICATION #superfamily giantin
SUMMARY         #length 3225 #molecular-weight 372240 #checksum 5760
Query Match      84.9%; Score 45; DB 2; Length 3225;
Best Local Similarity 75.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 251 AEQNOIL 258
||||:|
Qy 88 AEQRDDIL 95

RESULT          7
ENTRY           A56539 #type complete
TITLE           giantin - human
ALTERNATE_NAMES macrogolgin
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change
                22-Jun-1999
ACCESSIONS      A56539; S37536
REFERENCE        A56539
#authors        Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.;
#journal         Griffiths, G.; Renz, M.
#journal         Mol. Cell. Biol. (1994) 14:2564-2576
#title          Molecular genetic analyses of a 376-kilodalton Golgi complex
#title          membrane protein (giantin).
#cross-references MUID:94187728
#accession      A56539
##molecule_type mRNA
##residues      1-3259 ##label SEE
##cross-references EMBL:X75304; NID:g405714; PIDN:CAA53052.1;
                PID:g405715
GENETICS
#gene           GDB:GOLG1; GCP: GCP371
##cross-references GDB:454958
#map_position   3q13.31-3q13.31
CLASSIFICATION #superfamily giantin
KEYWORDS        coiled coil; Golgi apparatus; heptad repeat; transmembrane
                protein
FEATURE
3238-3254       #domain transmembrane #status predicted #label TMN
SUMMARY         #length 3259 #molecular-weight 376075 #checksum 4495
Query Match      84.9%; Score 45; DB 1; Length 3259;
Best Local Similarity 75.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 285 AEQNOIL 292
||||:|
Qy 88 AEQRDDIL 95

RESULT          8
ENTRY           B60176 #type complete
TITLE           hypothetical protein (Rv2744c region) - Mycobacterium

```

```

tuberculosis
#formal_name Mycobacterium tuberculosis
#formal_name Mycobacterium tuberculosis
17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
20-Sep-1999
ACCESSIONS      B60176
REFERENCE        A60176
#authors        O'Connor, S.P.; Rumschlag, H.S.; Mayer, L.W.
#journal         Res. Microbiol. (1990) 141:407-423
#title          Nucleotide sequence of the gene encoding the 35-kDa protein
#title          of Mycobacterium tuberculosis.
#cross-references MUID:90378140
#accession      B60176
##status        translation not shown
##molecule_type DNA
##residues      1-271 ##label OAC
##cross-references GB:M69187; NID:g149904; PID:g293240
COMMENT          This is the hypothetical translation of a sequence that was not
                  reported as a coding sequence in the complete genome. It overlaps
                  and is in a different reading frame from Rv2744c.
CLASSIFICATION #superfamily Mycobacterium tuberculosis hypothetical protein
                  (Rv2744c region)
SUMMARY         #length 271 #molecular-weight 29675 #checksum 3998
Query Match      83.0%; Score 44; DB 2; Length 271;
Best Local Similarity 62.5%; Pred. No. 3.01e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 212 AEQRHEVL 219
||||:|
Qy 88 AEQRDDIL 95

RESULT          9
ENTRY           I55184 #type complete
TITLE           repA protein Escherichia coli
ALTERNATE_NAMES #formal_name Escherichia coli
ORGANISM        07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change
DATE            02-Jul-1996
ACCESSIONS      I55184; S34178
REFERENCE        I55184
#authors        Llanes, C.; Gabant, P.; Couturier, M.; Michel-Briand, Y.
#journal         J. Bacteriol. (1993) 76:3403-3407
#title          Cloning and characterization of the Inc A/C plasmid RAI
#title          replicon.
#accession      I55184
##status        preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues      1-288 ##label RES
##cross-references EMBL:X73674; NID:g313284; PID:g313285
GENETICS
#gene           repA
SUMMARY         #length 288 #molecular-weight 32992 #checksum 3765
Query Match      83.0%; Score 44; DB 2; Length 288;
Best Local Similarity 62.5%; Pred. No. 3.01e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 110 AGOREEIL 117
||||:|
Qy 88 AEQRDDIL 95

RESULT          10
ENTRY           S50675 #type complete
TITLE           pre-mRNA splicing helicase BRR2 - yeast (Saccharomyces
                  cerevisiae)
ALTERNATE_NAMES DNA repair protein RAD24; protein YER172c
ORGANISM        #formal_name Saccharomyces cerevisiae
DATE            28-May-1993 #sequence_revision 24-Feb-1995 #text_change
                20-Sep-1999
ACCESSIONS      S50675; S30856; JC4657
REFERENCE        S50428
#authors        Dietrich, F.S.

```

```

#submission submitted to the EMBL Data Library, December 1994
#description the sequence of S. cerevisiae cosmids 9163 and 9132.
#accession S50675
##molecule_type DNA
##residues 1-2163 ##label D1E
##cross-references EMBL:U18922; NID:G603405; PIDN:AAB4699.1;
PID:G603413; MIPS:YER172C

REFERENCE
#S30812
#authors Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.;
Kemp, C.; Wei, Y.; Taylor, P.; Nakahara, K.; Roberts, D.;
Davis, R.W.
#submission submitted to the EMBL Data Library, February 1993
#accession S30856
##molecule_type DNA
##residues 1-169 ##label MUL
##cross-references EMBL:L11229

REFERENCE
#JC4657
#authors Zhu, Y.B.; Prakash, L.; Prakash, S.
#journal Chinese Biochem. J. (1995) 11:541-550
#title Molecular cloning and sequencing of DNA repair gene RAD24.
#accession JC4657
##molecule_type DNA
##residues 1-260, 'HCQT', 262, 'KT', 265 ##label ZHU
GENETICS
#gene SCD:RR2; RAD24
##cross-references SGD:S0000974; MIPS:YER172C
#map_position 5R
CLASSIFICATION
#superfamily unassigned DEAD/H box helicases; DEAD/H box
helicase homology
#ATP; DNA repair; nucleus; P-loop; pre-mRNA splicing
KEYWORDS
#domain DEAD/H box helicase homology #label DEAD\
#region nucleotide-binding motif A (P-loop)\
#region nucleotide-binding motif B\
#region DSX motif
#length 2163 #molecular-weight 246183 #checksum 6675
SUMMARY
Query Match 83.0%; Score 44; DB 2; Length 2163;
Best Local Similarity 71.4%; Pred. No. 3.01e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 2025 EERDIL 2031
QY 89 EORDIL 95

RESULT 11
ENTRY K1RTA #type complete
TITLE Ig kappa chain C region (allele a) - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 18-Aug-1982 #sequence_revision 10-Sep-1982 #text_change
22-Jun-1999
ACCESSION A02118
REFERENCE A93901
#authors Sheppard, H.W.; Gutman, G.A.
#journal Proc. Natl. Acad. Sci. U.S.A. (1981) 78:7064-7068
#title Allelic forms of rat kappa chain genes: evidence for strong
selection at the level of nucleotide sequence.
#cross-references MUID:82082587
#accession A02118
##molecule_type DNA
##residues 1-106 ##label SHE
##cross-references GB:J02574; GB:J00745; NID:G204820; PIDN:AAA41411.1;
PID:G204821
##experimental_source strain DA
COMPLEX
#An immunoglobulin heterotetramer subunit consists of two
identical light (kappa or lambda) and two identical heavy
(alpha, delta, epsilon, gamma, or mu) chains usually
stabilized by interchain disulfide bonds. In some cases,
such as IgA and IgM, the subunits associate into larger
oligomers.
#superfamily immunoglobulin C region; immunoglobulin homology
heterotetramer
KEYWORDS

#submission submitted to the EMBL Data Library, December 1994
#description the sequence of S. cerevisiae cosmids 9163 and 9132.
#accession S50675
##molecule_type DNA
##residues 1-2163 ##label D1E
##cross-references EMBL:U18922; NID:G603405; PIDN:AAB4699.1;
PID:G603413; MIPS:YER172C

REFERENCE
#S30812
#authors Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.;
Kemp, C.; Wei, Y.; Taylor, P.; Nakahara, K.; Roberts, D.;
Davis, R.W.
#submission submitted to the EMBL Data Library, February 1993
#accession S30856
##molecule_type DNA
##residues 1-169 ##label MUL
##cross-references EMBL:L11229

REFERENCE
#JC4657
#authors Zhu, Y.B.; Prakash, L.; Prakash, S.
#journal Chinese Biochem. J. (1995) 11:541-550
#title Molecular cloning and sequencing of DNA repair gene RAD24.
#accession JC4657
##molecule_type DNA
##residues 1-260, 'HCQT', 262, 'KT', 265 ##label ZHU
GENETICS
#gene SCD:RR2; RAD24
##cross-references SGD:S0000974; MIPS:YER172C
#map_position 5R
CLASSIFICATION
#superfamily unassigned DEAD/H box helicases; DEAD/H box
helicase homology
#ATP; DNA repair; nucleus; P-loop; pre-mRNA splicing
KEYWORDS
#domain DEAD/H box helicase homology #label DEAD\
#region nucleotide-binding motif A (P-loop)\
#region nucleotide-binding motif B\
#region DSX motif
#length 2163 #molecular-weight 246183 #checksum 6675
SUMMARY
Query Match 83.0%; Score 44; DB 2; Length 2163;
Best Local Similarity 71.4%; Pred. No. 3.01e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 2025 EERDIL 2031
QY 89 EORDIL 95

RESULT 11
ENTRY K1RTA #type complete
TITLE Ig kappa chain C region (allele a) - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 18-Aug-1982 #sequence_revision 10-Sep-1982 #text_change
22-Jun-1999
ACCESSION A02118
REFERENCE A93901
#authors Sheppard, H.W.; Gutman, G.A.
#journal Proc. Natl. Acad. Sci. U.S.A. (1981) 78:7064-7068
#title Allelic forms of rat kappa chain genes: evidence for strong
selection at the level of nucleotide sequence.
#cross-references MUID:82082587
#accession A02118
##molecule_type DNA
##residues 1-106 ##label SHE
##cross-references GB:J02574; GB:J00745; NID:G204820; PIDN:AAA41411.1;
PID:G204821
##experimental_source strain DA
COMPLEX
#An immunoglobulin heterotetramer subunit consists of two
identical light (kappa or lambda) and two identical heavy
(alpha, delta, epsilon, gamma, or mu) chains usually
stabilized by interchain disulfide bonds. In some cases,
such as IgA and IgM, the subunits associate into larger
oligomers.
#superfamily immunoglobulin C region; immunoglobulin homology
heterotetramer
KEYWORDS

#domain immunoglobulin homology #label IMM
#disulfide_bonds #status predicted\
#disulfide_bonds interchain (to heavy chain) #status
predicted
#length 106 #molecular-weight 11732 #checksum 4035
SUMMARY
Query Match 81.1%; Score 43; DB 1; Length 106;
Best Local Similarity 62.5%; Pred. No. 4.85e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 45 SEORDGVL 52
QY 88 AEQRDDIL 95

RESULT 12
ENTRY I44001 #type complete
TITLE nef protein - human immunodeficiency virus type 1 (strain
YU-2)
ALTERNATE_NAMES 3'-orf protein; orf-F protein
ORGANISM #formal_name human immunodeficiency virus type 1, HIV-1
#note host Homo sapiens (man)
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
23-Feb-1997
ACCESSION I44001
REFERENCE A44001
#authors Li, Y.; Hui, H.; Burgess, C.J.; Price, R.W.; Sharp, P.M.;
Hahn, B.H.; Shaw, G.M.
#journal J. Virol. (1992) 66:6587-6600
#title Complete nucleotide sequence, genome organization, and
biological properties of human immunodeficiency virus type
1 in vivo: evidence for limited defectiveness and
complementation.
#cross-references MUID:93021387
#accession I44001
##molecule_type DNA
##residues 1-214 ##label LIY
##cross-references GB:M93258
GENETICS
#gene nef; 3'-orf; orf-F
CLASSIFICATION #superfamily AIDS nef protein
KEYWORDS AIDS; immunodeficiency
SUMMARY
#length 214 #molecular-weight 24532 #checksum 5715
Query Match 81.1%; Score 43; DB 1; Length 214;
Best Local Similarity 62.5%; Pred. No. 4.85e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 113 SOQRDIL 120
QY 88 AEQRDDIL 95

RESULT 13
ENTRY ASLJBR #type complete
TITLE nef protein - human immunodeficiency virus type 1 (isolate
BR)
ALTERNATE_NAMES 3'-orf protein; orf-F protein
ORGANISM #formal_name human immunodeficiency virus type 1, HIV-1
#note host Homo sapiens (man)
DATE 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change
16-Jul-1999
ACCESSION D31667; S21993; S21995; S21997; S21999; S21991
REFERENCE A94389
#authors Anand, R.; Thayer, R.; Srinivasan, A.; Nayyar, S.; Gardner,
M.; Luciw, P.; Dandekar, S.
#journal Virology (1989) 168:79-89
#title Biological and molecular characterization of human
immunodeficiency virus (HIV-1-BR) from the brain of a
patient with progressive dementia.
#cross-references MUID:89085613
#accession D31667

```

```

##molecule_type DNA
##residues 1-218 ##label ANA
##cross-references GB:M21098; NID:g326426; PIDN:AAA44222.1; PID:g326431
REFERENCE S21990
#authors Steuler, H.; Storch-Hagenlocher, B.; Wildemann, B.; Hacke, W.
#submission Submitted to the EMBL Data Library, July 1991
#description Distinct populations of HIV-1 in blood and cerebrospinal
#fluid as determined by direct sequencing.
#accession S21993
##molecule_type DNA
##residues 1-7 ##label STE
##cross-references EMBL:X61358; NID:g60177; PIDN:CAA43629.1;
PID:g584027; EMBL:X61357; NID:g60175; PID:g584026
#accession S21995
##molecule_type DNA
##residues 1-7 ##label ST2
##cross-references EMBL:X61355; NID:g60179; PIDN:CAA43623.1;
PID:g584028; EMBL:X61357; NID:g60175; PID:g584026
#accession S21997
##molecule_type DNA
##residues 1-7 ##label ST3
##cross-references EMBL:X61356; NID:g60181; PIDN:CAA43625.1;
PID:g1129140; EMBL:X61357; NID:g60175; PID:g584026
#accession S21999
##molecule_type DNA
##residues 1-7 ##label ST4
##cross-references EMBL:X61359; NID:g60182; PIDN:CAA43631.1; PID:g584030
GENETICS
#gene nef; 3'-orf; orf-F
#superfamily AIDS nef protein
#AIDS: immunodeficiency
#length 218 #molecular-weight 25032 #checksum 9211
CLASSIFICATION
KEYWORDS
SUMMARY
Query Match 81.1%; Score 43; DB 1; Length 218;
Best Local Similarity 62.5%; Pred. No. 4.85e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 115 SQRDIL 122
|||:|
QY 88 AEQRDIL 95

RESULT 14
ENTRY QOCVP2 #type complete
TITLE coat protein, potato yellow mosaic virus (isolate Venezuela)
ORGANISM #formal_name potato yellow mosaic virus
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
23-Jul-1999
ACCESSIONS JU0367
REFERENCE JU0362
#authors Coutts, R.H.A.; Coffin, R.S.; Roberts, E.J.F.; Hamilton,
W.D.O.
#journal J. Gen. Virol. (1991) 72:1515-1520
#title The nucleotide sequence of the infectious cloned DNA
#components of potato yellow mosaic virus.
#cross-references M101311403
#accession JU0367
##molecule_type DNA
##residues 1-251 ##label COU
##cross-references GB:D00940; NID:g222458; PIDN:BAA00779.1;
PID:d1001242; PID:g222460
GENETICS
#map_position segment A
CLASSIFICATION #superfamily cassava latent virus coat protein
KEYWORDS coat protein
SUMMARY #length 251 #molecular-weight 29266 #checksum 3568
Query Match 81.1%; Score 43; DB 1; Length 251;
Best Local Similarity 85.7%; Pred. No. 4.85e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 73 EQRDIL 79
|||

```

```

QY 89 EQRDIL 95

RESULT 15
ENTRY C72589 #type complete
TITLE hypothetical protein APE1183 - Aeropyrum pernix (strain K1)
ORGANISM #formal_name Aeropyrum pernix
DATE 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
20-Aug-1999
ACCESSIONS C72589
REFERENCE A72450
#authors Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.;
Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba,
S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai,
Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura,
Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
#journal DNA Res. (1999) 6:83-101
#title Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
#cross-references M10199310339
#accession C72589
##status preliminary
##molecule_type DNA
##residues 1-309 ##label KAW
##cross-references DDBJ:AP000061; NID:g5104821; PIDN:BAA80169.1;
PID:d1043955; PID:g5104855
##experimental_source strain K1
GENETICS
#gene APE1183
SUMMARY #length 309 #molecular-weight 36033 #checksum 5836
Query Match 81.1%; Score 43; DB 2; Length 309;
Best Local Similarity 62.5%; Pred. No. 4.85e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 205 AEQRNEFL 212
|||:|
QY 88 AEQRDIL 95

```

Search completed: Wed May 10 12:50:26 2000
Job time : 8 secs.

Matches	6;	Conservative	1;	Mismatches	0;	Indels	0;	Gaps	0;
---------	----	--------------	----	------------	----	--------	----	------	----

RESULT	3	STANDARD	PRT	795 AA.
ID	P5CS_MOUSE			
AC	Q9ZL10			
AD	15-FEB-2000 (Rel. 39, Created)			
DT	15-FEB-2000 (Rel. 39, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	DELTA 1-PYRROLINE-5-CARBOXYLATE SYNTHETASE (P5CS) [INCLUDES: GLUTAMATE			
DE	5-KINASE (EC 2.7.2.11) (GAMMA-GLUTAMYL KINASE) (GK): GAMMA-GLUTAMYL			
DE	DE PHOSPHATE REDUCTASE (GPR) (EC 1.2.1.41) (GLUTAMATE-5-SEMIALDEHYDE			
DE	DE DHYDROGENASE) (GLUTAMATE-5-SEMIALDEHYDE DEHYDROGENASE)].			
GN	PYCS OR P5CS.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			

RP SEQUENCE FROM N.A.
RX MEDLINE; 99156365.
RA Hu C.A., Lin W.-W., Obie C., Valle D.;

RL J. Biol. Chem. 274:6754-6762(1999).

-! CATALYTIC ACTIVITY: L-GLUTAMATE 5-SEMIALDEHYDE + ORTHOPHOSPHATE +
NADP(+) = L-GAMMA-GLUTAMYL 5-PHOSPHATE + NADPH.
-! PATHWAY: FIRST AND SECOND STEPS IN PROLINE BIOSYNTHESIS PATHWAY.
-! SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE GLUTAMATE 5-
KINASE FAMILY.
-! SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GAMMA-
GLUTAMYL PHOSPHATE REDUCTASE FAMILY.

```

CC CC This SWISS-PROT entry is copyright. It is produced through a collabora
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstatio
CC CC the European Bioinformatics Institute. There are no restrictions on
CC CC use by non-profit institutions as long as its content is in no
CC CC modified and this statement is not removed. Usage by and for commerc
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announc
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC EMBL; AF056573; AAD17517.1; -.
CC DR PROSITE; PS00902; GLUTAMATE_5_KINASE; 1.
CC DR PROSITE; PS01223; PROA; 1
CC DR Proline biosynthesis; Multifunctional enzyme; Oxidoreductase; NADP;
CC KW Transferase; Kinase.
CC KW DOMAIN 1 361 GLUTAMATE 5-KINASE.
CC FT DOMAIN 362 795 GAMMA-GLUTAMYL PHOSPHATE REDUCTASE.
CC FT SEQUENCE 795 AA: 87296 MW: E3FCB9B334B69706 CRC64;
CC
CC Query Match 90.6%; Score 48; DB 1; Length 795;
CC Best Local Similarity 50.0%; Pred. No. 1.14e+00;
CC Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps

```

D8		397 TDQREIL 1404	#						
OY		:::::							
		88 AEQRDIL 95							
RESULT	4		#						
ID	PDR_BURCE	STANDARD;					PRT;	321 AA.	
AC	P33164;								
DT	01-OCT-1993	(Rel. 27,							
DT	01-OCT-1993	(Rel. 27, Last sequence update)							
DE	15-JUL-1998	(Rel. 36, Last annotation update)							
DT	PHYHALATE DIOXYGENASE REDUCTASE (EC 1.-.-.-)	(PDR).							

OS Burkholderia cepacia (Pseudomonas cepacia).
 OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
 OC Burkholderia.
 RN [1]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS), AND PARTIAL SEQUENCE.
 RC STRAIN=DS01.
 RX MEDLINE; 93088078.
 RA Correll C.C., Balle C.J., Ballou D.P., Ludwig M.L.;
 RT "Phthalate dioxygenase reductase: a modular structure for electron
 transfer from pyridine nucleotides to [2Fe-2S].";
 RL Science 258:1604-1610(1992).
 CC -!- FUNCTION: COMPONENT OF THE ELECTRON TRANSFER CHAIN INVOLVED IN
 CC PYRIDINE NUCLEOTIDE-DEPENDENT DIHYDROXYLATION OF PHTHALATE.
 CC UTILIZES FMN TO MEDIATE ELECTRON TRANSFER FROM THE TWO-ELECTRON
 CC DONOR, NADH, TO THE ONE-ELECTRON ACCEPTOR, (2FE-2S).
 CC -!- COFACTOR: FMN.
 CC -!- SUBUNIT: MONOMER.
 CC -!- SIMILARITY: IN THE C-TERMINAL, BELONGS TO THE 2FE2S PLANT-TYPE
 CC FERREDOXIN FAMILY.
 CC -!- SIMILARITY: BELONGS TO THE PDR/VANB FAMILY.
 CC -!- CAUTION: PART OF THIS SEQUENCE WAS DETERMINED BY X-RAY, ALANINE IS
 CC INDICATED IN POSITIONS WHERE NO SIDE CHAIN COULD BE OBSERVED.
 DR PIR; A44230; A44230.
 DR PDB; 2PIA; 15-APR-93.
 DR PRINTS; PR00409; PHDIOXRDTSF.
 DR PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
 DR PFAM; PF00111; fer2; 1.
 DR PFAM; PF00175; oxidoreductase; 1.
 DR Oxidoreductase; Flavoprotein; FMN; NAD; Iron-sulfur; 3D-structure;
 KW Electron transport.
 FT NP_BIND 1 102 FMN.
 FT NP_BIND 112 226 NAD.
 FT METAL 272 272 IRON-SULFUR (2FE-2S).
 FT METAL 277 277 IRON-SULFUR (2FE-2S).
 FT METAL 280 280 IRON-SULFUR (2FE-2S).
 FT METAL 308 308 IRON-SULFUR (2FE-2S).
 FT UNSURE 61 73
 FT UNSURE 164 168
 FT UNSURE 224 243
 FT UNSURE 256 261
 FT UNSURE 281 300
 FT TURN 3 7
 FT STRAND 9 20
 FT TURN 21 22
 FT STRAND 23 29
 FT TURN 31 32
 FT TURN 41 42
 FT STRAND 44 48
 FT TURN 50 51
 FT STRAND 54 58
 FT TURN 63 64
 FT STRAND 68 74
 FT TURN 77 78
 FT HELIX 82 88
 FT TURN 92 93
 FT STRAND 95 98
 FT TURN 102 102
 FT STRAND 109 110
 FT TURN 113 119
 FT STRAND 120 136
 FT HELIX 140 147
 FT TURN 150 152
 FT TURN 154 155
 FT TURN 156 160
 FT TURN 162 167
 FT STRAND 168 172
 FT TURN 174 175
 FT TURN 177 178
 FT HELIX 183 187
 FT TURN 192 193
 FT STRAND 194 199
 FT HELIX 202 211
 FT TURN 212 214

FT TURN 217 218
 FT STRAND 220 223
 FT STRAND 238 242
 FT TURN 243 245
 FT STRAND 248 251
 FT TURN 253 254
 FT HELIX 257 263
 FT TURN 264 265
 FT TURN 279 280
 FT STRAND 281 286
 FT STRAND 289 290
 FT TURN 298 303
 FT STRAND 304 306
 FT TURN 307 309
 FT STRAND 311 312
 FT STRAND 316 319
 SQ SEQUENCE 321 AA; 34785 MW; 78B1FD6D47E2AE39 CRC64;
 Query Match 88.7%; Score 47; DB 1; Length 321;
 Best Local Similarity 62.5%; Pred. No. 2.02e+00;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Db 289 ADHRDDVL 296
 QY 88 AEQRDDIL 95
 I:::|::|
 RESULT 5
 ID CCA_METH STANDARD; PRT: 454 AA.
 AC 026684;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE TRNA NUCLEOTIDYLTRANSFERASE (EC 2.7.7.25) (TRNA ADENYLYLTRANSFERASE)
 DE (TRNA CCA-PYROPHOSPHORYLASE) (CCA-ADDING ENZYME).
 GN CCA OR MTH584.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
 OC Methanobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DELTA H;
 RX MEDLINE; 98037514.
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Nollong J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 deltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -!- FUNCTION: THIS ENZYME CARRIES OUT SYNTHESIS OF THE TRNA CCA
 CC TERMINUS.
 CC -!- CATALYTIC ACTIVITY: ATP + TRNA(N) = PYROPHOSPHATE + TRNA(N+1).
 CC -!- SIMILARITY: BELONGS TO THE TRNA NUCLEOTIDYLTRANSFERASE / POLY(A)
 CC POLYMERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE000840; AAB85090.1;
 KW Transferase; Nucleotidyltransferase; RNA-binding; trna processing.
 SQ SEQUENCE 454 AA; 51345 MW; 5E79C1A3DE392A47 CRC64;
 Query Match 88.7%; Score 47; DB 1; Length 454;
 Best Local Similarity 85.7%; Pred. No. 2.02e+00;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 152 EQRDVL 158
 QY 89 EQRDIL 95

RESULT 6
 ID TP2B_MOUSE STANDARD; PRT; 1612 AA.
 AC Q64511;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
 GN TOP2B.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE-BRAIN;
 RA Miyake M., Adachi N., Kikuchi A.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
 BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
 MAKES DOUBLE-STRAND BREAKS.
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 OF DOUBLE-STRANDED DNA.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
 CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
 NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
 RELAX ONLY NEGATIVE SUPERCOILS.
 CC -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D38046; BAA07236.1;
 DR HSP; P06786; LBGW.
 DR MGD; MGI:98791; TOP2B.
 DR PRINTS; PR00418; TP12FAMILY.
 DR PRINTS; PR00615; CCAATSUBUNTA.
 DR PRINTS; PR01158; TOPISMRASEII.
 DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
 DR PFAM; PF00204; DNA_topoisoi; 1.
 KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
 FT NP_BIND 170 175 ATP (POTENTIAL).
 FT ACT_SITE 814 814 DNA CLEAVAGE (BY SIMILARITY).
 SQ SEQUENCE 1612 AA; 181863 MW; 331B9300651308C4 CRC64;
 Query Match 84.9%; Score 45; DB 1; Length 1612;
 Best Local Similarity 62.5%; Pred. No. 6.10e+00;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1189 AQERDIL 1196
 QY 88 AEQRDIL 95

RESULT 7
 ID TP2B_CRILLO STANDARD; PRT; 1612 AA.
 AC Q64599;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
 GN TOP2B.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE-BRAIN;
 RA Miyake M., Adachi N., Kikuchi A.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
 BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
 MAKES DOUBLE-STRAND BREAKS.
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 OF DOUBLE-STRANDED DNA.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
 CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
 NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
 RELAX ONLY NEGATIVE SUPERCOILS.
 CC -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D38046; BAA07236.1;
 DR HSP; P06786; LBGW.
 DR MGD; MGI:98791; TOP2B.
 DR PRINTS; PR00418; TP12FAMILY.
 DR PRINTS; PR00615; CCAATSUBUNTA.
 DR PRINTS; PR01158; TOPISMRASEII.
 DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
 DR PFAM; PF00204; DNA_topoisoi; 1.
 KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
 FT NP_BIND 170 175 ATP (POTENTIAL).
 FT ACT_SITE 814 814 DNA CLEAVAGE (BY SIMILARITY).
 SQ SEQUENCE 1612 AA; 181863 MW; 331B9300651308C4 CRC64;
 Query Match 84.9%; Score 45; DB 1; Length 1612;
 Best Local Similarity 62.5%; Pred. No. 6.10e+00;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1189 AQERDIL 1196
 QY 88 AEQRDIL 95

RESULT 8
 ID BRR2_YEAST STANDARD; PRT; 2163 AA.
 AC P32639;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PRE-MRNA SPLICING HELICASE BRR2 (EC 3.6.1.-).
 GN BRR2 OR RSS1 OR YER172C OR SYGP-ORF66.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
 RA Aviles E., Berne A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
 RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
 RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,

OS Cricetus longicaudatus (Long-tailed hamster) (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LUNG;
 RX MEDLINE; 96085121.
 RA Derudder S., Frey S., Delaporte C., Jacquemin-Sablon A.;
 RT "Cloning and characterization of full-length cDNAs coding for the DNA
 topoisomerase II beta from Chinese hamster lung cells sensitive and
 resistant 9-OH-ellipticine";
 RL Biochim. Biophys. Acta 1264:178-182(1995).
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
 BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
 MAKES DOUBLE-STRAND BREAKS.
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 OF DOUBLE-STRANDED DNA.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
 CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
 NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
 RELAX ONLY NEGATIVE SUPERCOILS.
 CC -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X86455; CAA60173.1;
 DR HSP; P06786; LBGW.
 DR PRINTS; PR00418; TP12FAMILY.
 DR PRINTS; PR00615; CCAATSUBUNTA.
 DR PRINTS; PR01158; TOPISMRASEII.
 DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
 DR PFAM; PF00204; DNA_topoisoi; 1.
 KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
 FT NP_BIND 170 175 ATP (POTENTIAL).
 FT ACT_SITE 814 814 DNA CLEAVAGE (BY SIMILARITY).
 SQ SEQUENCE 1612 AA; 182074 MW; COLD6FC40620FC68 CRC64;
 Query Match 84.9%; Score 45; DB 1; Length 1612;
 Best Local Similarity 62.5%; Pred. No. 6.10e+00;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1189 AQERDIL 1196
 QY 88 AEQRDIL 95

RESULT 9
 ID BRR2_YEAST STANDARD; PRT; 2163 AA.
 AC P32639;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PRE-MRNA SPLICING HELICASE BRR2 (EC 3.6.1.-).
 GN BRR2 OR RSS1 OR YER172C OR SYGP-ORF66.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
 RA Aviles E., Berne A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
 RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
 RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,

RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-169 FROM N.A.
RA Mulligan J.T., Dietrich F.S., Hennessey K.M., Sehl P., Komp C.,
RA Wei Y., Taylor P., Nakahara K., Roberts D., Davis R.W.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RC STRAIN-DBY473;
RX MEDLINE: 96304576.
RA Noble S.M., Guthrie C.;
RT "Identification of novel genes required for yeast pre-mRNA splicing
by means of cold-sensitive mutations.";
RL Genetics 143:67-80(1996).
CC -!- FUNCTION: REQUIRED FOR PRE-MRNA SPLICING.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SKI2 SUBFAMILY OF HELICASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: U18922; AAB64699.1; -.
DR PIR: S30856; S30856.
DR SGD: L0003100; BR2.
DR PFAM: PF00270; DEAD. 2.
KW Helicase; ATP-binding; Nuclear protein; mRNA processing.
FT NP_BIND 79 86
FT NP_BIND 521 528 ATP (POTENTIAL).
FT SITE 634 637 DEH BOX.
FT SITE 634 637
SQ SEQUENCE 2163 AA; 246183 MW; DFAF7E3B7168D944 CRC64;

Query Match 83.08; Score 44; DB 1; Length 2163;
Best Local Similarity 71.4; Pred. No. 1.05e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 2025 EERDEIL 2031
QY 89 EQRDDIL 95
:|||||

RESULT 9
ID KACA-RAT STANDARD; PRT; 106 AA.
AC P01836;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG KAPPA CHAIN C REGION, A ALLELE.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DA;
RX MEDLINE: 82082587.
RA Sheppard H.W., Gutman G.A.;
RT "Allelic forms of rat kappa chain genes: evidence for strong
selection at the level of nucleotide sequence.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:7064-7068(1981).
DR PIR: A02118; K1RTA.
DR PROSITE: PS00290; IG_MHC; 1.
DR PFAM: PF00047; ig; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
FT DISULFID 26 86
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
SQ SEQUENCE 106 AA; 11732 MW; B7E120D9700DD566 CRC64;

Query Match 81.1%; Score 43; DB 1; Length 106;
Best Local Similarity 62.5%; Pred. No. 1.78e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 45 SEQRDVL 52
QY 88 AEQRDIL 95
:|||||

RESULT 10
ID NEF_HV1Y2 STANDARD; PRT; 214 AA.
AC P35959;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE NEGATIVE FACTOR (F-PROTEIN) (27 KD PROTEIN) (3'ORF).
GN NEF.
OS Human immunodeficiency virus type 1 (YU-2 isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93021387.
RA Li Y., Hui H., Burgess C.J., Price R.W., Sharp P.M., Hahn B.H.,
RA Shaw G.M.;
RT "Complete nucleotide sequence, genome organization, and biological
properties of human immunodeficiency virus type 1 in vivo: evidence
for limited defectiveness and complementation.";
RL J. Virol. 66:6587-6600(1992).
CC -!- FUNCTION: NEF HAS GTPASE, GTP-BINDING AND AUTOPHOSPHORYLATING
ACTIVITIES, IT SEEM TO DOWN REGULATE THE CD4(T4) ANTIGEN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: M93258; -; NOT_ANNOTATED_CDS.
DR PIR: I44001; I44001.
DR HSSP: P03406; IEPN.
DR PFAM: PF00469; F-protein; 1.
KW AIDS; Myristate; GTP-binding.
FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
SQ SEQUENCE 214 AA; 24532 MW; E188D43D7B084D04 CRC64;

Query Match 81.1%; Score 43; DB 1; Length 214;
Best Local Similarity 62.5%; Pred. No. 1.78e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 113 SQQRDIL 120
QY 88 AEQRDIL 95
:|||||

RESULT 11
ID NEF_HV1B STANDARD; PRT; 218 AA.
AC P12479;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NEGATIVE FACTOR (F-PROTEIN) (27 KD PROTEIN) (3'ORF).
GN NEF.
OS Human immunodeficiency virus type 1 (1BR isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89085613.
RA Anand R., Thayer R., Srinivasan A., Nayyar S., Gardner M., Luciw P.,
RA Dandekar S.;
RT "Biological and molecular characterization of human immunodeficiency
virus (HIV-1BR) from the brain of a patient with progressive

```
RT dementia.";
RL Virology 168:79-89(1989).
CC -!- FUNCTION: NEF HAS GTPASE, GTP-BINDING AND AUTOPHOSPHORYLATING
CC ACTIVITIES. IT SEEM TO DOWN REGULATE THE CD4(T4) ANTIGEN.
CC -!- MISCELLANEOUS: THIS VIRUS IS CYTOPATHICALLY ACTIVE AND WAS
CC HARVESTED FROM THE BRAIN TISSUE OF A NEUROLOGICAL AIDS PATIENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M21098; AAA44222.1; -
CC PIR; D31667; ASLJBR.
CC HSP; P03406; ILEN.
CC HIV; M21098; NEFSRVA.
CC PFAM; PF00469; F-protein; 2.
CC AIDS; Myristate; GTP-binding.
CC LIPID 2
CC MYRISTATE (BY SIMILARITY).
CC SEQUENCE 218 AA; 25032 MW; FC4DAEA1045C460E CRC64;
CC -----
CC Query Match 81.1%; Score 43; DB 1; Length 218;
CC Best Local Similarity 82.5%; Pred. No. 1.78e+01; Indels 0; Gaps 0;
CC Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC Db 115 SOORDBIL 122
CC QY 88 AEQRDIL 95
CC -----
RT RESULT 12
ID COAT_PYMV STANDARD; PRT; 251 AA.
AC P27255;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE COAT PROTEIN.
OS Potato yellow mosaic virus (isolate Venezuela).
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91311403.
RA Coutts R.H.A., Coffin R.S., Roberts E.J.F., Hamilton W.D.O.;
RT "The nucleotide sequence of the infectious cloned DNA components of
RT potato yellow mosaic virus.";
RL J. Gen. Virol. 72:1515-1520(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D00940; BAA00779.1; -
CC PIR; J00367; QOCVP2.
CC PFAM; PF00844; Gemini_coat; 1.
CC PRINTS; PR00223; GEMCOATBR1.
CC PRINTS; PR00224; GEMCOATAR1.
CC Coat protein.
CC SEQUENCE 251 AA; 29266 MW; 98939DD842AE3A55 CRC64;
CC -----
CC Query Match 81.1%; Score 43; DB 1; Length 251;
CC Best Local Similarity 85.7%; Pred. No. 1.78e+01; Indels 0; Gaps 0;
CC Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CC -----
CC Db 73. EQRDIL 79
CC QY 89 EQRDIL 95
CC -----
RT RESULT 13
ID RIR2_TREPA STANDARD; PRT; 351 AA.
AC O83092;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE BETA CHAIN (EC 1.17.4.1)
DE (RIBONUCLEOTIDE REDUCTASE).
GN NRDB OR TP0053
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NICHOLS;
RX MEDLINE: 98332770.
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -!- FUNCTION: CATALYZES THE BIOSYNTHESIS OF DEOXYRIBONUCLEOTIDES FROM
CC THE CORRESPONDING RIBONUCLEOTIDES, PRECURSORS THAT ARE NECESSARY
CC FOR DNA SYNTHESIS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2'-DEOXYRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
CC THIOREDOXIN + H(2)O = RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
CC THIOREDOXIN.
CC -!- COFACTOR: CONTAINS TWO IRON IONS (BY SIMILARITY).
CC -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC SMALL CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001190; AAC65049.1; -
CC TIGR; TP0053; -
CC PROSITE; PS00368; RIBORED_SMALL; 1.
CC PFAM; PF00268; ribonuc_red; 1.
CC Oxidoreductase; DNA replication; Iron.
CC METAL 94 IRON 1 (BY SIMILARITY).
CC METAL 124 IRON 1 AND 2 (BY SIMILARITY).
CC METAL 127 IRON 1 (BY SIMILARITY).
CC METAL 191 IRON 2 (BY SIMILARITY).
CC METAL 225 IRON 2 (BY SIMILARITY).
CC METAL 228 IRON 2 (BY SIMILARITY).
CC ACT_SITE 131 BY SIMILARITY.
CC SEQUENCE 351 AA; 41198 MW; 5B373A7FCBFF049A CRC64;
CC -----
CC Query Match 81.1%; Score 43; DB 1; Length 351;
CC Best Local Similarity 82.5%; Pred. No. 1.78e+01;
CC Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC Db 141 PERNDIL 148
CC QY 88 AEQRDIL 95
CC -----
RT RESULT 14
ID YF20_HAEIN STANDARD; PRT; 355 AA.
```

AC P44240;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL PROTEIN H11520.
GN H11520.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20.
RX MEDLINE: 95350630.
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Klenney A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- SIMILARITY: TO B.SUBTILIS XKDT AND YQBT AND TO E.COLI YMFP.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U32827; AAC23162.1; -
DR TIGR; H11520; -
KW Hypothetical protein.
SQ SEQUENCE 355 AA; 38452 MW; D708248CB5871895 CRC64;
Query Match 81.1%; Score 43; DB 1; Length 355;
Best Local Similarity 57.1%; Pred. No. 1.78e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 229 DEREDIL 235
QY 89 EQRDDIL 95

RESULT 15
ID LAT.NOCIA STANDARD; PRT; 450 AA.
AC Q05174;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE L-LYSINE-EPSILON AMINOTRANSFERASE (EC 2.6.1.36) (L-LYSINE
DE AMINOTRANSFERASE) (LYSINE 6-AMINOTRANSFERASE).
GN LAT.
OS Nocardia lactamdurans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92011390.
RA Coque J., Liras P., Laiz L., Martin J.;
RT "A gene encoding lysine 6-aminotransferase, which forms the
RT beta-lactam precursor alpha-aminoadipic acid, is located in the
RT cluster of cephamycin biosynthetic genes in Nocardia lactamdurans.";
RL J. Bacteriol. 173:6258-6264(1991).
CC -!- CATALYTIC ACTIVITY: L-LYSINE + 2-OXOGLUTARATE = 2-AMINOADIPATE
CC 6-SEMIALDEHYDE + L-GLUTAMATE.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -!- PATHWAY: BETA-LACTAM BIOSYNTHESIS: FIRST STEP IN CEPHAMYCIN C

CC SYNTHESIS.
CC -!- SUBUNIT: ACTIVE AS EITHER A MONOMER OR A HOMOPOLYMER.
CC -!- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z21681; CAA79796.1; -
DR PIR; A38171; A38171.
DR PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
DR PFAM; PF00202; aminotra_3; 2.
KW Transférase; Aminotransferase; Antibiotic biosynthesis;
KW Pyridoxal phosphate.
FT BINDING 300 300 PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ SEQUENCE 450 AA; 48806 MW; C24915CAF2E93EE5 CRC64;
Query Match 81.1%; Score 43; DB 1; Length 450;
Best Local Similarity 62.5%; Pred. No. 1.78e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 396 TEORDAVL 403
QY 88 AEQRDDIL 95

Search completed: Wed May 10 12:45:44 2000
Job time : 97 secs.

